



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12P 23/00, 7/26, C12N 9/02, 1/20, 15/00, C07H 21/04, C07K 14/00		A1	(11) International Publication Number: WO 99/61652 (43) International Publication Date: 2 December 1999 (02.12.99)
(21) International Application Number: PCT/US99/10455 (22) International Filing Date: 21 May 1999 (21.05.99)		(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).	
(30) Priority Data: 60/086,460 22 May 1998 (22.05.98) US		Published <i>With international search report. Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i>	
(71) Applicant (for all designated States except US): UNIVERSITY OF MARYLAND [US/US]; Office of Technology Liaison, 4312 Knox Road, College Park, MD 20742 (US).			
(72) Inventor; and (75) Inventor/Applicant (for US only): CUNNINGHAM, Francis, X. [US/US]; 2727 Washington Avenue, Chevy Chase, MD 20815 (US).			
(74) Agents: GOLDHUSH, Douglas, H. et al.; Nikaido, Marmel- stein, Murray & Oram LLP, Metropolitan Square, Suite 330, G Street Lobby, 655 Fifteenth Street, N.W., Washington, DC 20005-5701 (US).			
(54) Title: CAROTENOID KETOLASE GENES AND GENE PRODUCTS, PRODUCTION OF KETOCAROTENOID AND METH- ODS OF MODIFYING CAROTENOIDUS USING THE GENES			
(57) Abstract			
<p>A purified nucleic acid sequence which encodes for a protein having ketolase enzyme activity and has the nucleic acid sequence of SEQ ID NO: 1 or 3, or has a sequence which encodes the amino acid sequence of SEQ ID NO: 2 or 4, as well as vectors and host cells containing them. Methods of use of the nucleic acid sequences to produce ketocarotenoid in host cells and methods of use of the nucleic acid sequences to modify the production of carotenoids in a host cell are included.</p>			

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republit of Korea	PT	Portugal		
CN	China	KZ	Kazakstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

**CAROTENOID KETOLASE GENES AND GENE PRODUCTS,
PRODUCTION OF KETOCAROTENOIDS AND METHODS OF
MODIFYING CAROTENOIDS USING THE GENES**

BACKGROUND OF THE INVENTION

5 Carotenoids are widely distributed natural pigments that are responsible for many of the yellow, orange and red colors seen in living organisms. They have important commercial uses as coloring agents in the food industry, as feed and food additives, in cosmetics and as provitamin A precursors.

10 The plant species *Adonis aestivalis* produces flowers with petals that are deep red in color and nearly black at the base of the petals due to the accumulation of ketocarotenoid and other carotenoid pigments (Neamtu et al., *Rev. Roum. Biochim.* 6:157, 1969). This pattern of carotenoid accumulation accounts for the common name of some varieties of this species: summer pheasant's eye.

15 Among the carotenoids identified in the petals of the red petal varieties of these various species is the ketocarotenoid astaxanthin (3,3'-dihydroxy-4,4'-diketo- β,β -carotene; see Figure 1). Various other ketocarotenoids (see Figure 1) including 3-hydroxyechinenone (3-hydroxy-4-keto- β,β -carotene), adonirubin (3-hydroxy-4,4'-diketo- β,β -carotene) adonixanthin (3,3'-dihydroxy-4-keto- β,β -carotene) and isozeaxanthin (4,4'-dihydroxy- β,β -carotene; see T.W. Goodwin, *The Biochemistry of the Carotenoids*, 20 vol I. Plants, 2nd edition, 1980, page 147) have also been reported. The latter compound is consistent with speculation that the 4-hydroxy may be an intermediate in the formation of the 4-keto group.

SUMMARY OF THE INVENTION

25 There is appreciable interest in the biological production of carotenoids, in particular the orange-colored ketocarotenoids such as astaxanthin and canthaxanthin (Figure 1), and in the modification of carotenoid composition. For this reason, an *A. aestivalis* flower cDNA library was constructed and screened for cDNAs encoding enzymes (hereinafter referred to as "ketolases" although the specific biochemical activity has not yet been established) involved in the conversion of β -carotene into 30 orange compounds with absorption properties similar to those exhibited by common ketocarotenoids such as canthaxanthin (Figure 1). Two distinctly different *Adonis aestivalis* cDNAs were obtained from among a number of cDNAs that were selected on this basis.

- 2 -

Thus, a first aspect of the present invention is a purified nucleic acid sequence which encodes for a protein having ketolase enzyme activity and has the nucleic acid sequence of SEQ ID NO: 1 or 3.

5 The invention also includes a purified nucleic acid sequence which encodes for a protein having ketolase enzyme activity and having the amino acid sequence of SEQ ID NO: 2 or 4.

The invention also includes vectors which comprise any portion of the nucleic acid sequences listed above, and host cells transformed with such vectors.

10 Another aspect of the present invention is a method of producing a ketocarotenoid in a host cell, the method comprising

15 inserting into the host cell a vector comprising a heterologous nucleic acid sequence which encodes for a protein having ketolase enzyme activity and comprises (1) SEQ ID NO: 1 or 3 or (2) a sequence which encodes the amino acid sequence of SEQ ID NO: 2 or 4, wherein the heterologous nucleic acid sequence is operably linked to a promoter; and

expressing the heterologous nucleic acid sequence, thereby producing the ketolase enzyme.

20 Another subject of the present invention is a method of modifying the production of carotenoids in a host cell, relative to an untransformed host cell, the method comprising

25 inserting into a host cell which already produces carotenoids a vector comprising a heterologous nucleic acid sequence which encodes for a protein having ketolase enzyme activity and comprises (1) SEQ ID NO: 1 or 3 or (2) a sequence which encodes the amino acid sequence of SEQ ID NO: 2 or 4, wherein the heterologous nucleic acid sequence is operably linked to a promoter; and

expressing the heterologous nucleic acid sequence in the host cell to modify the production of the carotenoids in the host cell, relative to an untransformed host cell.

BRIEF DESCRIPTION OF THE DRAWINGS

30 A more complete appreciation of the invention and many of the attendant advantages thereof will be readily obtained as the same becomes better understood by

- 3 -

reference to the following detailed description when considered in connection with the accompanying drawings.

Figure 1 illustrates structures and biochemical routes leading from β -carotene to various of the ketocarotenoids referred to in the text. Conversion of β -carotene to astaxanthin 5 by a hydroxylase enzyme (Hy) and a ketolase enzyme (keto) could proceed via any one or all of several possible routes depending on the order of the reactions.

Figure 2 illustrates the *beta* ring structure of β -carotene and various modifications of this parent ring that might be produced through the action of the products of the *A. aestivalis* ketolase cDNAs. Also shown is the structure of the *epsilon* ring, not found to be a 10 substrate for the *A. aestivalis* ketolases and present in carotenoids such as d-carotene, e-carotene, a-carotene and lutein.

Figure 3 illustrate results obtained with TLC (thin layer chromatography) separation of carotenoid pigments extracted from *E. coli* cultures, previously engineered to produce β -carotene, but that now also contain the *A. aestivalis* ketolase cDNAs and/or other 15 introduced genes and cDNAs. The Figure indicates the empty plasmid vector pBluescript SK- (SK-), the *Adonis aestivalis* ketolase 1 cDNA in this plasmid vector (Ad keto1), the *Haematococcus pluvialis* ketolase cDNA in this plasmid vector Hp keto), or the *Arabidopsis* β -carotene hydroxylase cDNA (At Ohase). Bands that were orange in color are shown here with a darker fill than those with a yellow color. Identities of 20 various bands are indicated to the right of the band.

Figure 4 illustrates the absorption spectrum of one of the orange carotenoids produced from β -carotene via the action of the *Adonis* ketolases and makes clear the similarity of the spectrum to that of canthaxanthin. Absorption spectra (in acetone) of β -carotene, canthaxanthin and an unknown orange product (orange band #1; the lower orange 25 band in the first lane of Figure 3) extracted from cultures after introduction of the *Adonis aestivalis* keto1 cDNA (SEQ ID NO: 1) in cells of *E. coli* that otherwise produce and accumulate β -carotene. The absorption spectrum of the unknown resembles that of canthaxanthin but the compound migrates to a position below echinenone on RP18

- 4 -

TLC plates developed with a mobile phase of methanol:acetone (1:1 by volume). The absorption spectrum of orange band #2 also is similar to that of canthaxanthin but it migrates more rapidly than canthaxanthin indicating that it is probably a more polar compound.

5 Figure 5 shows SEQ ID NO: 5 (the sequence shown in this Figure includes SEQ ID NO: 1 and also includes some of the flanking DNA from the adaptor DNA and the multiple cloning site (MCS) of the library cloning vector, which sequences are shown in bold).

Figure 6 shows SEQ ID NO: 6 (the sequence shown in this Figure includes SEQ ID NO: 2 and also includes a translation of amino acids resulting from the adaptor DNA and 10 the multiple cloning site (MCS) of the library cloning vector and the start codon from the plasmid vector pTrcHis, which sequences are shown in bold and capitalized).

Figure 7 shows SEQ ID NO: 7 (the sequence shown in this Figure includes SEQ ID NO: 3 and also includes some of the flanking DNA from the adaptor DNA and the multiple cloning site (MCS) of the library cloning vector, which sequences are shown in bold).

15 Figure 8 shows SEQ ID NO: 8 (the sequence shown in this Figure includes SEQ ID NO: 4 and also includes a translation of amino acids resulting from the adaptor DNA and the multiple cloning site (MCS) of the library cloning vector and the start codon from the plasmid vector, which sequences are shown in bold and capitalized).

Figure 9 shows a "Gap" alignment of the two Adonis ketolase sequences of the 20 invention. A truncated version of SEQ ID NO: 1 is shown in this Figure for comparative purposes, and is designated SEQ ID NO: 9. The percentage identity was calculated to be 91.107.

Figure 10 shows a "Gap" alignment of SEQ ID NO: 2 and 4. The following results were found:

25 Gap weight: 12 average match: 2.912
 Length weight: 4 average mismatch: -2.003

- 5 -

Quality: 1440 length: 307
Ratio: 4.691 gaps: 0
percent similarity: 92.182 percent identity: 90.228

Figure 11 shows a comparison between SEQ ID NO: 2 and the *Arabidopsis thaliana* β -5 carotene hydroxylase enzyme (GenBank U58919) (SEQ ID NO: 10).

Figure 12A shows gDNA (SEQ ID NO: 11) immediately upstream of the cDNA of SEQ ID NO: 3. The sequence was obtained from a PCR product generated using the GenomeWalker kit of Clontech Laboratories, Inc. (1020 East Meadow Circle, Palo Alto, CA 94303-4230) and nested primers specific to the ketolases of *Adonis aestivalis* 10 (cagaatcggtctgttctatttagttctcc (SEQ ID NO: 17) and caatttgaggaatatcaagggttccttgtc (SEQ ID NO: 18)). The termination codon upstream of and in-frame with initiation codon (TAA at positions 204-206) is shown in bold. Initiation codon (ATG) is also shown in bold.

Figure 12B (SEQ ID NO: 12) indicates that the full length polypeptide of SEQ ID NO: 15 4 begins with the amino acids MAA (shown in bold) immediately preceding the ketolase sequence shown in Figure 8. A similar MAA amino acid sequence immediately preceding SEQ ID NO: 1 is also expected.

Figure 13 shows an alignment of SEQ ID NO: 2, SEQ ID NO: 12, an *Arabidopsis* β -20 carotene hydroxylase enzyme (predicted product of GenBank U58919) (SEQ ID NO: 13), a putative second *Arabidopsis* hydroxylase predicted by genomic DNA sequence (GenBank AB025606; the exon/intron junctions were chosen with reference to the product of the *Arabidopsis* β -carotene hydroxylase cDNA u58919) (SEQ ID NO: 14), and two *Capsicum annuum* β -carotene hydroxylases (predicted products of GenBank 25 Y09722 and Y09225) (SEQ ID NO: 15 and 16).

DESCRIPTION OF THE PREFERRED EMBODIMENTS

The present invention is directed to a purified nucleic acid sequence which

- 6 -

encodes for a protein having ketolase enzyme activity and has the nucleic acid sequence of SEQ ID NO: 1 or 3.

The invention also includes a purified nucleic acid sequence which encodes for a protein having ketolase enzyme activity and having the amino acid sequence of SEQ 5 ID NO: 2 or 4.

Two different but closely-related nucleic acids have been isolated. The sequences of the longest example of each are presented herein. Sequencing which has subsequently been conducted of upstream genomic DNA indicates that SEQ ID NO: 3 lacks bases encoding the first three amino acids (MAA; see Figure 12). Likely, 10 this is also the case for SEQ ID NO: 1, but the upstream genomic sequences have not yet been obtained for this nucleic acid.

The two different Adonis ketolases denoted in SEQ ID NO: 1 and 3 are similar in sequence, sharing about 91% identity, as determined by the Gap program discussed below (see Figure 9). The predicted amino acid sequences of the enzymes denoted in 15 SEQ ID NO: 2 and 4 share about 92% similarity and about 90% identity, also as determined by the Gap program (see Figure 10).

Therefore, it is clear that certain modifications of SEQ ID NO: 1 or 3 or SEQ ID NO: 2 or 4 can take place without destroying the activity of the enzyme. Note also that certain truncated versions of the cDNAs of SEQ ID NO: 1 or 3 were found to be 20 functional (i.e., these cDNAs retained the property of causing the conversion of β -carotene to orange compounds). Also, the *Arabidopsis* β -carotene hydroxylase (GenBank U58919), aligned with the ketolase SEQ ID NO: 2 in Figure 11, retains catalytic function when truncated to yield a polypeptide that lacks the first 129 amino acids (Sun et al., 1996). From the alignment in Figure 11, therefore, this would suggest 25 that the two ketolases of the invention retain catalytic activity after truncation to remove bases encoding the first 132 amino acids.

Thus, the present invention is intended to include those ketolase nucleic acid and amino acid sequences in which substitutions, deletions, additions or other modifications have taken place, as compared to SEQ ID NO: 1 or 3 or SEQ ID NO: 2 30 or 4, without destroying the activity of the ketolase enzyme. Preferably, the substitutions, deletions, additions or other modifications take place at those positions which already show dissimilarity between the present sequences. For SEQ ID NO: 1,

- 7 -

as shown in Figure 9, these positions are as follows: positions 7, 20, 23, 35, 53, 63, 65, 67, 76, 78, 85, 86, 91, 107, 109-111, 135, 140, 144, 146, 160, 168, 217, 219, 241, 249, 254, 256, 271, 291, 296, 349, 389, 400, 406, 431, 448, 449, 460, 471, 499, 530, 589, 619, 643, 653, 654, 667, 679, 709, 731, 742, 784, 787, 836, 871, 883, 896, 911, 919, 928, 930, 939, 943, 967, 969, 978, 979, 982, 988, 995, 1005, 1006, 1012-1014, 1017, 1019-1021, 1023, 1025, 1049, 1050, 1054, 1060-1068, 1070-1073, 1075, 1094, 1100, 1101, 1106, 1107, 1109 and 1111-1176. For SEQ ID NO: 3, as shown in Figure 9, these positions are as follows: positions 7, 20, 23, 35, 53, 63, 65, 67, 76, 78, 85, 86, 91, 107, 109-111, 135, 140, 144, 146, 160, 168, 217, 219, 241, 249, 254, 256, 271, 291, 296, 349, 389, 400, 406, 431, 448, 449, 460, 471, 499, 530, 589, 619, 643, 653, 654, 667, 679, 709, 731, 742, 784, 787, 836, 871, 883, 896, 911, 919, 928, 930, 939, 943, 966, 967, 970, 979, 980, 983, 989, 996, 1006, 1007, 1013-1015, 1018, 1020-1022, 1024, 1026, 1050, 1051, 1055, 1062-1065, 1067, 1086, 1092, 1093, 1098, 1099, 1101 and 1103-1112.

For SEQ ID NO: 2 and 4, as shown in Figure 10, the following amino acids can be substituted or deleted, or additions or other modifications can be made, without destroying the activity of the ketolase enzyme: positions 7, 8, 12, 18, 21, 22, 25, 26, 36, 37, 45, 47-49, 56, 73, 83, 85, 97, 99, 130, 144, 150, 157, 166, 218, 244, 279, 299 and 304. Therefore, the present invention also intends to cover amino acid sequences where such changes have been made.

In each case, nucleic acid and amino acid sequence similarity and identity is measured using sequence analysis software, for example, the Sequence Analysis, Gap, or BestFit software packages of the Genetics Computer Group (University of Wisconsin Biotechnology Center, 1710 University Avenue, Madison, Wisconsin 53705), MEGAlign (DNAStar, Inc., 1228 S. Park St., Madison, Wisconsin 53715), or MacVector (Oxford Molecular Group, 2105 S. Bascom Avenue, Suite 200, Campbell, California 95008). Such software uses algorithms to match similar sequences by assigning degrees of identity to various substitutions, deletions, and other modifications, and includes detailed instructions as to useful parameters, etc., such that those of routine skill in the art can easily compare sequence similarities and identities. An example of a useful algorithm in this regard is the algorithm of Needleman and Wunsch, which is used in the Gap program discussed above. This program finds the alignment of two complete

sequences that maximizes the number of matches and minimizes the number of gaps. Another useful algorithm is the algorithm of Smith and Waterman, which is used in the BestFit program discussed above. This program creates an optimal alignment of the best segment of similarity between two sequences. Optimal alignments are found by 5 inserting gaps to maximize the number of matches using the local homology algorithm of Smith and Waterman.

Conservative (i.e. similar) substitutions typically include substitutions within the following groups: glycine and alanine; valine, isoleucine and leucine; aspartic acid, glutamic acid, asparagine and glutamine; serine and threonine; lysine and arginine; and 10 phenylalanine and tyrosine. Substitutions may also be made on the basis of conserved hydrophobicity or hydrophilicity (see Kyte and Doolittle, *J. Mol. Biol.* **157**: 105-132 (1982)), or on the basis of the ability to assume similar polypeptide secondary structure (see Chou and Fasman, *Adv. Enzymol.* **47**: 45-148 (1978)).

If comparison is made between nucleotide sequences, preferably the length of 15 comparison sequences is at least 50 nucleotides, more preferably at least 60 nucleotides, at least 75 nucleotides or at least 100 nucleotides. It is most preferred if comparison is made between the nucleic acid sequences encoding the enzyme coding regions necessary for enzyme activity. If comparison is made between amino acid sequences, preferably the length of comparison is at least 20 amino acids, more 20 preferably at least 30 amino acids, at least 40 amino acids or at least 50 amino acids. It is most preferred if comparison is made between the amino acid sequences in the enzyme coding regions necessary for enzyme activity.

While the two different Adonis ketolase enzymes of the present invention are similar in sequence, previously-described bacterial (Misawa et al., 1995), cyanobacterial 25 (Fernandez-Gonzalez et al., 1997), and green algal (*Haematococcus pluvialis*; Lotan et al., 1995; Kajiwara et al., 1995) β -carotene ketolase enzymes bear little resemblance to the Adonis ketolases, although certain histidine motifs and features of the predicted secondary structure are common to the polypeptides predicted by both groups (Cunningham and Gantt, 1998).

30 The present invention also includes vectors containing the nucleic acids of the invention. Suitable vectors according to the present invention comprise a gene encoding a ketolase enzyme as described above, wherein the gene is operably linked

- 9 -

to a suitable promoter. Suitable promoters for the vector can be constructed using techniques well known in the art (see, for example, Sambrook et al., Molecular Cloning A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1989; Ausubel et al., Current Protocols in Molecular Biology, Greene Publishing and Wiley 5 Interscience, New York, 1991). Suitable vectors for eukaryotic expression in plants are described in Fray et al., (1995; *Plant J.* 8:693-701) and Misawa et al, (1994; *Plant J.* 6:481-489). Suitable vectors for prokaryotic expression include pACYC184, pUC119, and pBR322 (available from New England BioLabs, Beverly, MA) and pTrcHis (Invitrogen) and pET28 (Novagen) and derivatives thereof. The vectors of the present 10 invention can additionally contain regulatory elements such as promoters, repressors, selectable markers such as antibiotic resistance genes, etc., the construction of which is very well known in the art.

The genes encoding the ketolase enzymes as described above, when cloned into a suitable expression vector, can be used to overexpress these enzymes in a host 15 cell expression system or to inhibit the expression of these enzymes. For example, a vector containing a gene of the invention may be used to increase the amount of ketocarotenoids in an organism and thereby alter the nutritional or commercial value or pharmacology of the organism. A vector containing a gene of the invention may also be used to modify the carotenoid production in an organism.

20 Therefore, the present invention includes a method of producing a ketocarotenoid in a host cell, the method comprising

inserting into the host cell a vector comprising a heterologous nucleic acid sequence which encodes for a protein having ketolase enzyme activity and comprises 25 (1) SEQ ID NO: 1 or 3 or (2) a sequence which encodes the amino acid sequence of SEQ ID NO: 2 or 4, wherein the heterologous nucleic acid sequence is operably linked to a promoter; and

expressing the heterologous nucleic acid sequence, thereby producing the ketocarotenoid.

The present invention also includes a method of modifying the production of 30 carotenoids in a host cell, relative to an untransformed host cell, the method comprising

inserting into a host cell which already produces carotenoids a vector comprising a heterologous nucleic acid sequence which encodes for a protein having

- 10 -

ketolase enzyme activity and comprises (1) SEQ ID NO: 1 or 3 or (2) a sequence which encodes the amino acid sequence of SEQ ID NO: 2 or 4, wherein the heterologous nucleic acid sequence is operably linked to a promoter; and

expressing the heterologous nucleic acid sequence in the host cell to

5 modify the production of the carotenoids in the host cell, relative to an untransformed host cell.

The term "modifying the production" means that the amount of carotenoids produced can be enhanced, reduced, or left the same, as compared to an untransformed host cell. In accordance with one embodiment of the present invention,

10 the make-up of the carotenoids (i.e., the type of carotenoids produced) is changed *vis a vis* each other, and this change in make-up may result in either a net gain, net loss, or no net change in the amount of carotenoids produced in the cell. In accordance with another embodiment of the present invention, the production or the biochemical activity of the carotenoids (or the enzymes which catalyze their formation) is enhanced by the

15 insertion of the ketolase enzyme-encoding nucleic acid. In yet another embodiment of the invention, the production or the biochemical activity of the carotenoids (or the enzymes which catalyze their formation) may be reduced or inhibited by a number of different approaches available to those skilled in the art, including but not limited to such methodologies or approaches as anti-sense (e.g., Gray et al. (1992), *Plant Mol. Biol.* 19:69-87), ribozymes (e.g., Wegener et al (1994) *Mol. Gen. Genet.* 1994 Nov 15;245(4):465-470), co-suppression (e.g. Fray et al. (1993) *Plant Mol. Biol.* 22:589-602), targeted disruption of the gene (e.g., Schaefer et al. *Plant J.* 11:1195-1206, 1997), intracellular antibodies (e.g., see Rondon et al. (1997) *Annu. Rev. Microbiol.* 51:257-283) or whatever other approaches rely on the knowledge or

20 25 availability of the nucleic acid sequences of the invention, or the enzymes encoded thereby.

Host systems according to the present invention preferably comprise any organism which is capable of producing carotenoids, or which already produces carotenoids. Such organisms include plants, algae, certain bacteria, cyanobacteria and

30 other photosynthetic bacteria. Transformation of these hosts with vectors according to the present invention can be done using standard techniques. See, for example, Sambrook et al., Molecular Cloning A Laboratory Manual, Cold Spring Harbor

- 11 -

Laboratory, Cold Spring Harbor, NY, 1989; Ausubel et al., Current Protocols in Molecular Biology, Greene Publishing and Wiley Interscience, New York, 1991.

Alternatively, transgenic organisms can be constructed which include the nucleic acid sequences of the present invention. The incorporation of these sequences can 5 allow the controlling of carotenoid biosynthesis, content, or composition in the host cell. These transgenic systems can be constructed to incorporate sequences which allow for the overexpression of the various nucleic acid sequences of the present invention. Transgenic systems can also be constructed which allow for the underexpression of the 10 various nucleic acid sequences of the present invention. Such systems may contain anti-sense expression of the nucleic acid sequences of the present invention. Such anti-sense expression would result in the accumulation of the substrates of the enzyme 15 encoded by the sense strand.

Having generally described this invention, a further understanding can be obtained by reference to certain specific examples which are provided herein for 15 purposes of illustration only and are not intended to be limiting unless otherwise specified.

EXAMPLE 1

Isolation of plant cDNAs that convert b-carotene into compounds with ketocarotenoid-like spectra

20 A flower cDNA library from the plant *Adonis aestivalis* was introduced into a strain of *Escherichia coli* engineered to accumulate the yellow carotenoid pigment β -carotene (see Cunningham et al., *Plant Cell* 8:1613-26, 1996). This strain of *E. coli* normally forms yellow colonies when cultures are spread on a solid agar growth medium. Ketocarotenoids that are derived from b-carotene, such as echinenone and 25 canthaxanthin (Figure 1), are, in contrast, orange to orange-red in color. Colonies that were orange rather than yellow in color were visually selected, and the DNA sequences of the *Adonis aestivalis* cDNAs within the plasmid vectors contained in these colonies were ascertained. Two distinct cDNAs were obtained from analysis of cDNA inserts in plasmids obtained from approximately 10 selected colonies. The DNA sequences of 30 these two ketolase cDNAs are presented herein.

The products produced by the ketolases of the invention which have been

expressed in a β -carotene-accumulating strain of *Escherichia coli* have not yet been identified. As many as 5 or 6 different colored bands, in addition to the substrate β -carotene, may readily be discerned by C_{18} TLC separation (see Figure 3). To provide appropriate standards to assist in identification, an *H. pluvialis* ketolase and an 5 *Arabidopsis* β -carotene hydroxylase were separately introduced into the β -carotene- accumulating *E. coli* to produce echinenone (3-keto- β,β -carotene) and canthaxanthin (3,3'-diketo- β,β -carotene) or β -cryptoxanthin (4-hydroxy- β,β -carotene) and zeaxanthin (4,4'-dihydroxy- β,β -carotene). None of the compounds formed in the presence of the 10 ketolases of the invention (no difference was observed in products formed in the presence of the two different nucleic acid sequences of the invention) both migrate in the TLC system and have the absorption spectrum expected for echinenone, canthaxanthin, β -cryptoxanthin, or zeaxanthin. Two of the colored TLC bands produced in the presence of the *Adonis* ketolase cDNAs are orange in color. Orange band #1 has an absorption spectrum similar to that of canthaxanthin (see Figure 4) but migrates 15 in a position that indicates a polarity intermediate to echinenone and β -carotene. Orange band #2 also has an absorption spectrum like that of canthaxanthin but migrates in a position that indicates a polarity intermediate to canthaxanthin and zeaxanthin (see Figure 3). The absorption spectra and TLC results suggest that the two orange products could be desaturated at the 3-4 positions of both rings (3,4,- 20 didehydro; see Figure 2). Orange band #1 (see Figure 3) might then be 3,4,3',4'-tetradehydro- β,β -carotene. To substantially affect the absorption spectrum of the substrate β -carotene, any modifications very likely involve a carbon that lies in conjugation with the conjugated chain of carbon-carbon double bonds that constitute the chromophore (Goodwin, 1980; The Biochemistry of the Carotenoids, volume I; 2nd 25 edition, Chapman and Hall). For the spectra obtained, only the carbons at the number 4 position of the two rings appear to be plausible locations for modification. The multitude and TLC migrations of the yellow and orange products produced from the symmetrical β -carotene, however, also indicates that the enzymes of the invention carry out more than a single type of reaction. The apparent homology of the ketolases of the 30 invention to the *Arabidopsis* β -carotene hydroxylase would suggest that compounds with a hydroxyl at the 3 and/or 4 positions of one or both rings are another possible outcome (see Figure 2). In fact, such compounds have been identified in *Adonis* (see

- 13 -

above), and it has long been conjectured that a hydroxyl at position 4 is an intermediate in the formation of the 4-keto (e.g. crustaxanthin, a 3,3',4,4' tetrahydroxy carotenoid that might be a precursor for astaxanthin in the exoskeleton of the lobster). The histidine motifs and secondary structure in common to the hydroxylase and ketolase enzymes

5 are characteristics of a large group of di-iron oxygenases whose members also include examples of desaturases (J. Shanklin, 1998, *Ann. Rev. Plant Physiol. Plant Mol. Biol.*), therefore a 3-4 desaturation (and/or perhaps a 2-3 desaturation in one or more of the yellow compounds) would also seem a plausible outcome.

To summarize the results of this example for the Adonis ketolases of the 10 invention, a number of different carotenoids, including two with ketocarotenoid-like spectra, are produced from β -carotene via the action of the products of either of the two different nucleic acids of the invention. These orange compounds appear to be the major products. Truncation and fusion of the cDNAs to a stronger promoter in the 15 vector pTrcHis (Invitrogen) was detrimental to growth of *E. coli* but did result in improved yield of the most polar orange product (orange band #2 in Figure 3). Introduction of a cyanobacterial ferredoxin did not change the yield or relative amounts 20 of the various products. Without being bound by theory, it may be that the ketocarotenoids produced in flower petals of Adonis actually include the as yet unidentified orange compounds that are produced in *E. coli* using the nucleic acids of the invention.

EXAMPLE 2

Substrate specificity of the Adonis ketolases

Carotenoids with ϵ rings are common in plants. The ϵ ring differs from the β ring only in the position of the double bond within the ring (Figure 2). The ϵ ring is reported 25 to be a poor substrate for the Arabidopsis β -carotene hydroxylase (Sun et al., 1996). The Adonis ketolase cDNAs were introduced into strains of *E. coli* engineered (Cunningham et al., 1996) to accumulate carotenoids with one or two ϵ rings (d-carotene and ϵ -carotene), or the acyclic carotenoid lycopene. TLC analysis of acetone extracts revealed that these carotenoids were not modified by the Adonis ketolases. 30 as indicated by a lack of any new products formed. Products produced in *E. coli* engineered to accumulate zeaxanthin (Sun et al., 1996) appeared to be the same as

- 14 -

for β -carotene accumulating cultures indicating that a 3-OH is likely to be one of the functional groups introduced to the β ring by the Adonis ketolases. The more polar orange band produced from β -carotene through the action of the Adonis ketolases (e.g., orange band 2 in Figure 3), therefore, could very well be 3,3'-dihydroxy-3,4,3',4'-
5 tetradehydro- β , β -carotene.

The references cited in the application, along with the following references, are incorporated by reference:

Bouvier F, et al. (1998) Xanthophyll biosynthesis: molecular and functional characterization of carotenoid hydroxylases from pepper fruits (*Capsicum annuum* L.).
10 Biochim Biophys Acta. 1391:320-8

Breitenbach J, et al. (1996) Expression in *Escherichia coli* and properties of the carotene ketolase from *Haematococcus pluvialis*. FEMS Microbiol Lett. 140:241-6

Cunningham FX Jr, Gantt E (1998) Genes and enzymes of carotenoid biosynthesis in plants. Ann Rev Plant Physiol Plant Mol Biol 49: 557-583

15 Fernandez-Gonzalez B, et al. (1997) A new type of asymmetrically acting beta-carotene ketolase is required for the synthesis of echinenone in the cyanobacterium *Synechocystis* sp. PCC 6803. J Biol Chem. 272:9728-33

Fraser PD, et al. (1997) In vitro characterization of astaxanthin biosynthetic enzymes. J Biol Chem. 1997272:6128-35

20 Fraser PD, et al. (1998) Enzymic confirmation of reactions involved in routes to astaxanthin formation, elucidated using a direct substrate in vitro assay. Eur J Biochem. 252:229-36

Harker M, et al. (1997) Biosynthesis of ketocarotenoids in transgenic cyanobacteria expressing the algal gene for beta-C-4-oxygenase, crtO. FEBS Lett. 404:129-34

- 15 -

Kajiwara S, et al. (1995) Isolation and functional identification of a novel cDNA for astaxanthin biosynthesis from *Haematococcus pluvialis*, and astaxanthin synthesis in *Escherichia coli*. *Plant Mol Biol*. 29:343-52

Lotan T, et al. (1995) Cloning and expression in *Escherichia coli* of the gene encoding 5 beta-C-4-oxygenase, that converts beta-carotene to the ketocarotenoid canthaxanthin in *Haematococcus pluvialis*. *FEBS Lett*. 364:125-8

Misawa N, et al. (1995) Canthaxanthin biosynthesis by the conversion of methylene to keto groups in a hydrocarbon beta-carotene by a single gene. *Biochem Biophys Res 10 Commun*.209:867-76

Misawa N, et al. (1995) Structure and functional analysis of a marine bacterial carotenoid biosynthesis gene cluster and astaxanthin biosynthetic pathway proposed at the gene level. *J Bacteriol*. 177:6575-84

Miura Y, et al. (1998) Production of the carotenoids lycopene, beta-carotene, and 15 astaxanthin in the food yeast *Candida utilis*. *Appl Environ Microbiol*. 64:1226-9

Shanklin J, et al. (1997) Mossbauer studies of alkane omega-hydroxylase: evidence for a diiron cluster in an integral-membrane enzyme. *Proc Natl Acad Sci U S A*. 94:2981-6

Shanklin J, Cahoon EB (1998) Desaturation and related modifications of fatty acids. *Ann Rev Plant Physiol Plant Mol Biol* 49: 611-641

20 Wang CW, et al. Engineered isoprenoid pathway enhances astaxanthin production in *Escherichia coli*. *Biotechnol Bioeng*. 1999 Jan 20;62(2):235-41.

I claim:

1. A method of producing a ketocarotenoid in a host cell, the method comprising
inserting into the host cell a vector comprising a heterologous nucleic acid
sequence which encodes for a protein having ketolase enzyme activity and has the
5 nucleic acid sequence of SEQ ID NO: 1 or 3, wherein the heterologous nucleic acid
sequence is operably linked to a promoter; and
expressing the heterologous nucleic acid sequence, thereby producing
the ketocarotenoid.
2. The method of claim 1, wherein the host cell is selected from the group
10 consisting of a bacterial cell, an algal cell and a plant cell.
3. A method of producing a ketocarotenoid in a host cell, the method comprising
inserting into the host cell a vector comprising a heterologous nucleic acid
sequence which encodes for a protein having ketolase enzyme activity and has a
sequence which encodes the amino acid sequence of SEQ ID NO: 2 or 4, wherein the
15 heterologous nucleic acid sequence is operably linked to a promoter; and
expressing the heterologous nucleic acid sequence, thereby producing
the ketocarotenoid.
4. The method of claim 3, wherein the host cell is selected from the group
consisting of a bacterial cell, an algal cell and a plant cell.
- 20 5. A method of modifying the production of carotenoids in a host cell, relative to an
untransformed host cell, the method comprising
inserting into a host cell which already produces carotenoids a vector
comprising a heterologous nucleic acid sequence which encodes for a protein having
ketolase enzyme activity and has the nucleic acid sequence of SEQ ID NO: 1 or 3,
25 wherein the heterologous nucleic acid sequence is operably linked to a promoter; and
expressing the heterologous nucleic acid sequence in the host cell to
modify the production of the carotenoids in the host cell, relative to an untransformed

host cell.

6. The method of claim 5, wherein the host cell is selected from the group consisting of a bacterial cell, an algal cell and a plant cell.

7. A method of modifying the production of carotenoids in a host cell, relative to an
5 untransformed host cell, the method comprising

inserting into a host cell which already produces carotenoids a vector comprising a heterologous nucleic acid sequence which encodes for a protein having ketolase enzyme activity and has a sequence which encodes the amino acid sequence of SEQ ID NO: 2 or 4, wherein the heterologous nucleic acid sequence is operably
10 linked to a promoter; and

expressing the heterologous nucleic acid sequence in the host cell to modify the production of the carotenoids in the host cell, relative to an untransformed host cell.

8. The method of claim 7, wherein the host cell is selected from the group
15 consisting of a bacterial cell, an algal cell and a plant cell.

9. A purified nucleic acid sequence which encodes for a protein having ketolase enzyme activity and has the nucleic acid sequence of SEQ ID NO: 1.

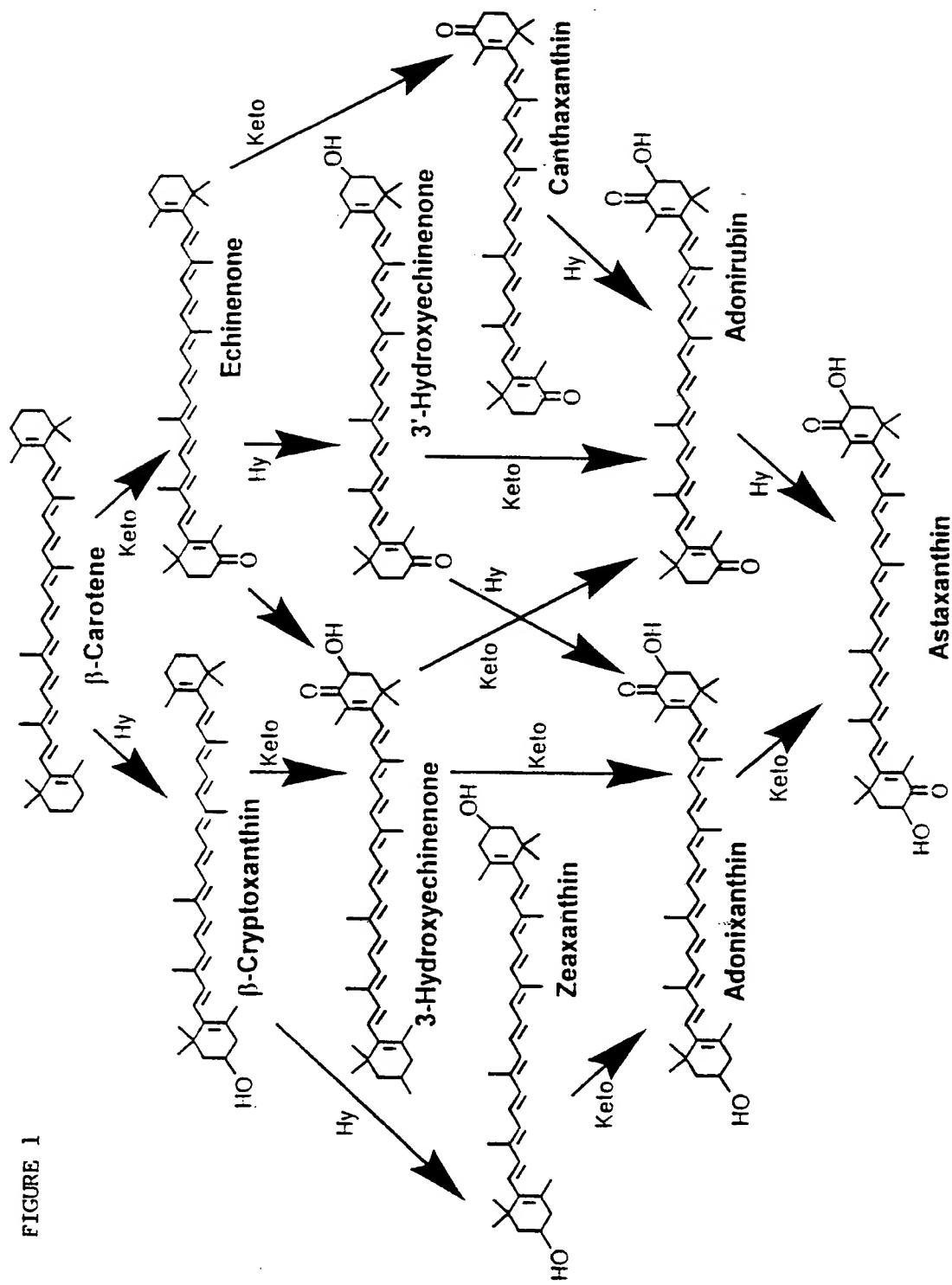
10. A purified nucleic acid sequence which encodes for a protein having ketolase enzyme activity and has the nucleic acid sequence of SEQ ID NO: 3.

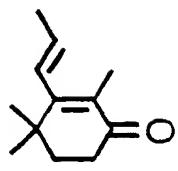
20 11. A purified nucleic acid sequence which encodes for a protein having ketolase enzyme activity and has a sequence which encodes the amino acid sequence of SEQ ID NO: 2.

12. A purified nucleic acid sequence which encodes for a protein having ketolase enzyme activity and has a sequence which encodes the amino acid sequence of SEQ
25 ID NO: 4.

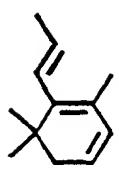
- 18 -

13. A vector which comprises the nucleic acid sequence of any one of claims 9-12, wherein the nucleic acid sequence is operably linked to a promoter.
14. A host cell which is transformed with the vector of claim 13.
15. The host cell of claim 14, wherein the host cell is selected from the group consisting of a bacterial cell, an algal cell and a plant cell.
16. The host cell of claim 14, wherein the host cell is a photosynthetic cell.
17. The host cell of claim 14, wherein the host cell contains a ketocarotenoid.
18. The host cell of claim 14, wherein the host cell contains modified levels of carotenoids, relative to an untransformed host cell.
19. A purified ketolase enzyme which is encoded by the amino acid sequence of SEQ ID NO: 2.
20. A purified ketolase enzyme which is encoded by the amino acid sequence of SEQ ID NO: 4.

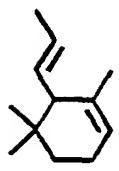




4-keto



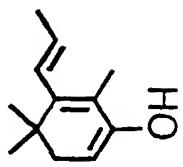
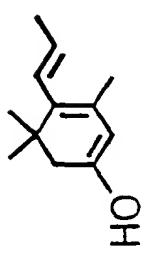
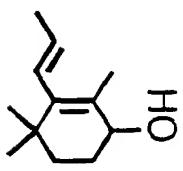
3,4-didehydro



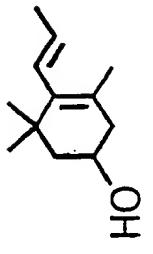
ε ring



β ring

3,4-didehydro
4-hydroxy3,4-didehydro
3-hydroxy

4-hydroxy



3-hydroxy

FIGURE 2

FIGURE 3

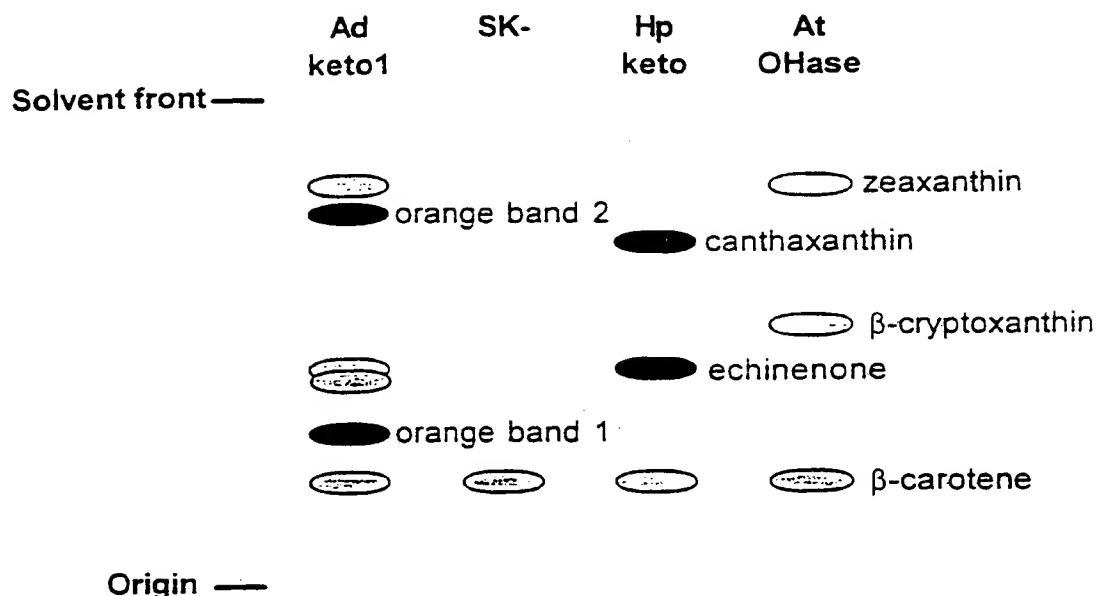


FIGURE 4

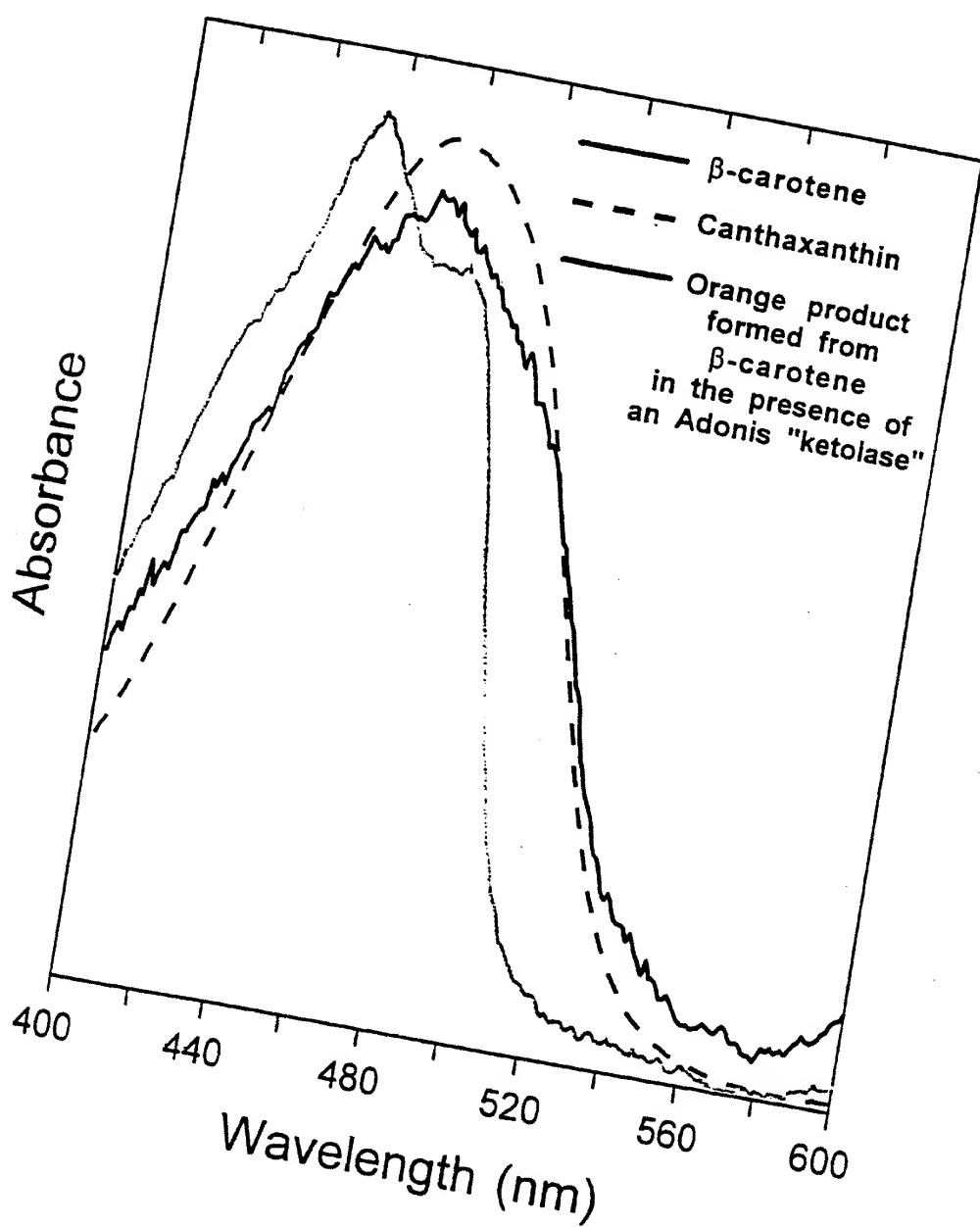


Figure 5 [SEQ ID NO: 5]

-23 ggg ctgcaggaat tcggcacgag
1 agcaatctca gtgttcagta caagttattc tttccacaag aatctcttgt
51 tgcactcaaa acaagacatt ctcacccgccc catgtttgct cttctctcca
101 gttgtgggggg agtcgcctat gagaaagaaa aagacacatc gtgctgcatg
151 tatctgctct gttgcagaga gaacaaggaa ccttgatatt cctcaaattg
201 aagaagagga agagaacgag gaagaactaa tagaacagac ggattctggc
251 ataattccata taaagaaaaac gctagggggg aaacaatcaa gacggtccac
301 tggctccatt gtgcaccccg tatcttgtct tgggatcctt tcaatgatcg
351 gaccctgcgtt ttacttcaag ttttcacggc taatggagtg tggagatatt
401 cctgtcgccag aaatggggat tacgtttgcc gcctttgttg ctgctgcgat
451 tggcacggaa ttttgcgtcag gatgggttca caaagaactc tggcacgatt
501 ctttgcgtta cattcacaag tctcaccata ggtcacgaaa aggccgcctc
551 gagttcaatg atgtgtttgc tattattaac gcgccttcctg ctattgctct
601 tatcaattat ggattctcaa atgaaggcct cttccttggc gcctgcgttg
651 gtaccggctt tggAACGACA gtctgtggca tggcttacat ttttcttcac
701 aatggcccttt cacaccgaag gttcccagta gggcttattg caaacgtccc
751 ttattttccac aagctggctg cagctcacca aatccatcac tcaggaaaat
801 ttcagggtt accatttggc ctgttccttg gaccccagga attggaagaa
851 gtaagacggag gcactgaaga attggagagg gtgatcagtc gtacagctaa
901 acgaacgc当地 tcatctacat gaatcaactc ttttacattt atgaggttt
951 agtttatcgg tggtacaagt cacacatttgc tgcgttgta gtaattcaaa
1001 gttaccatac tcttttttag aattttttt tgatgtatag gtcgcggagt
1051 tacgggttaca aaggccaaat ctattgtgt ggaattccat tattaaaaat
1101 aaaaatttgc当地 gtttgcgtt ttatctggc atcaatatca atatatatta
1151 attaaagc当地 aaaaaaaaaaaa aaaaaaa ctcgag

6/16

Figure 6 [SEQ ID NO: 6]

MGLQEFGTR

aisvfsts sys fhknlllhsk qdilnrc l1 fspvvvespm rkkkthraac
icsvaertrn ldipqieeee eneeelieqt dsgiihikkt lggkqsrrst
gsivapvscl gilsmigpav yfkfsrlmec gdipvaemgi tfaafvaaai
gteflsgwvn kelwhdslwy ihkshh rsk grfefndvfa iinalpaial
inygfsnegl lpgacfgtgl gttvcgmayi flhnglshrr fpvglianvp
yfhklaahq ihhsgkfqgv pfglflgpqe leevrggtee iervisrtak
rtqsst*

7/16

Figure 7 [SEQ ID NO: 7]

-23 ggg ctgcaggaat tcggcacgag
1 agcaattca gtgttcagtt caggttattc tttctacaag aatctttgt
51 tggactcaaa accaaatatt ctcaaaccgg catgcctgct attctctcca
101 gttgtgatca tgcgcctat gagaaagaaa aagaaacatg gtgatccatg
151 tatctgctcc gttgcagggaa gaacaaggaa ctttgatatt cctcaaattg
201 aagaagagga agagaatgtg gaagaactaa tagaacagac cgattctgac
251 atagtgcatat aaaaacactaaggaa aaacaatcaa aacggcccac
301 tggctccatt gtcgcacccg tatcttgcct tggatcctt tcaatgatg
351 gacctgctgt tttttcaag ttttacggc taatggaggg tggagatata
401 cctgttagcag aaatggggat tacgtttgcc acctttgttgc tgcgtctgt
451 tggcacggag tttttgtcag catgggttca caaagaactc tggcacgagt
501 ctttggta cattcacaag tctcaccatc ggtcacgaaa agggcgcttc
551 gagttcaatg atgtgtttgc tattattaac ggcgttcccg ctattgtct
601 tatcaattat ggattctcca atgaaggcct ctttcttggaa gcgtgttttgc
651 gtgtcggtct tggacaacaaca gtcgtggta tggcttacat ttttcttcc
701 aatggccat cacaccgaag tttccagta tggcttatttgc cgaacgtccc
751 ttatttccac aagctggctg cagctcacca aatacaccac tcaggaaaat
801 ttcagggtgt accatttggc ctgttttttgc gacccaaagga attggaaagaa
851 gtaagaggag gcactgaaga tttggagagg gtaatcagtc gtacaactaa
901 acgaacgcaa ccatctacccatc gaatcaattt ttttacatataaataagggttt
951 agtttatcgg ttttataaaaa tcacacatcc gtatcggtt agtaagtcaa
1001 agttaagata cttccttctt agaatatttt ttgatgtata ggtcgccggat
1051 atactgttac actattcggtt gtggaaattcc attataaaaaa aataaaaaaaa
1101 aaaaaaaaaaa aa ctcgag

8/16

Figure 8 [SEQ ID NO: 8]

MGLQEFGTR

aisvfssgys fyknillldsk pnilkppc11 fspvvimspm rkkkkhgdpc
icsvagrtrn ldipqieeee enveelieqt dssdivhikkt lggkqskrpt
gsivapvscl gilmigpav yfkfsrlmeg gdipvaemgi tfatfvaav
gteflsawvh kelwneslwy ihkshhersrk grfefndvfa iinalpaial
inygfsnegl lpgacfvgvl gttvcgmayi flhnglshrr fpvwlianvp
yfhklaahq ihsqkfqgv pfglflgpke leevrggtee lervisrttk
rtqpst*

9/16

Figure 9: Gap of SEQ ID NO: 9 and SEQ ID NO: 3

1 agcaatctcagtgttcagtacaagttattcttccacaagaatctcttgt 50
||||||| ||||||||| ||| ||||||||| ||||||||| |||||||||
1 agcaatttcagtgttcagttcaggttattcttctacaagaatctcttgt 50

51 tgcactcaaaacaagacattctcaaccgcccattttgtctttctcca 100
||| ||||||| | | ||||| | | ||||| | ||| |||||
51 tggactcaaaacccaaatattctcaaaccatgcctgtattctcca 100

101 gttgtggggaggcgccatgagaaagaaaaagacacatcggtgtgcattg 150
||||||| | | ||||| ||||| ||||| ||||| ||| | |||||
101 gttgtgatcatgtcgccatgagaaagaaaaagaaacatggtgatccatg 150

151 tatctgctttgcagagagaacaaggAACCTTgatattcctcaaattg 200
||||||| | | ||||| ||||| ||||| ||||| ||||| |||||
151 tatctgctccgttgcagggagaacaaggAACCTTgatattcctcaaattg 200

201 aagaagaggaaagagaacgaggaaagaactaatagaacagacggattctggc 250
||||||| ||||| | | ||||| ||||| ||||| ||||| |||
201 aagaagaggaaagagaatgtggaagaactaatagaacagacggattctgac 250

251 ataattcatataaagaaaacgctaggggggaaacaatcaaggacggtccac 300
||| | ||||| ||||| ||||| ||||| ||||| |||||
251 atagtgcataaaaagaaaacacttagggggaaacaatcaaaacggccac 300

301 tggctccatgtcgacccgtatctgtcttggatccttcaatgatcg 350
||||||| ||||| ||||| ||||| ||||| ||||| |||||
301 tggctccatgtcgacccgtatctgtcttggatccttcaatgattg 350

10/16

Figure 9 (cont.)

351 gacctgctgtttacttcaagtttacggctaattggagtggtggagatatt 400
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
351 gacctgctgtttacttcaagtttacggctaattggagggtggagatata 400

401 cctgtcgacaaaatgggattacgttgcgccttgttgctgctgcgt 450
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |
401 cctgttagcagaaaatgggattacgttgcaccccttggctgctgt 450

451 tggcacggaaattttgcaggatgggtcacaaaagaactctggcacgatt 500
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |
451 tggcacggagttttgcagcatgggtcacaaaagaactctggcacgagt 500

501 ctttgcgttacattcacaagtctcaccataggcacgaaaaggccgcttc 550
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
501 ctttgcgttacattcacaagtctcaccatcggtcacgaaaaggccgcttc 550

551 gagttcaatgatgtttgttattattaaacgcgttccgttattgtct 600
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
551 gagttcaatgatgtttgttattattaaacgcgttccgttattgtct 600

601 tatcaattatggattctcaaatttgcggcccttcgttgcgttttgcgtt 650
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
601 tatcaattatggattctcaaatttgcggcccttcgttgcgttgcgtt 650

651 gtaccggcttggaaacgacagtctgtggcatggcttacattttcttac 700
|| | ||||| ||||| ||||| ||||| ||||| ||||| |||||
651 gtgtcggtcttggaaacaacagtctgtggatggcttacattttcttac 700

11/16

Figure 9 (cont.)

701 aatggccttcacaccgaagggtcccagtagggcttattgcaaacgtccc 750
||||||| ||||||||| ||||||||| ||||||||| |||||||||
701 aatggcctatcacaccgaagggtcccagtagggcttattgcaaacgtccc 750

751 ttatttccacaagctggctgagctcaccaatccatcactcaggaaaat 800
||||||| ||||||||| ||||||||| ||| |||||||||
751 ttatttccacaagctggctgagctcaccaatacaccactcaggaaaat 800

801 ttcaagggttaccattggcctgttcctggaccccaggaatttggaaagaa 850
||||||| ||||||||| ||||||||| |||||||||
801 ttcaagggttaccattggcctgttcctggacccaaggaaatttggaaagaa 850

851 gtaagaggaggcactgaagaattggagaggggtatcagtcgtacagctaa 900
||||||| ||||||||| ||||||||| ||||||||| |||||
851 gtaagaggaggcactgaagagttggagagggtaatcagtcgtacaactaa 900

901 acgaacgcaatcatctacaTGAatcaactctttacatttatgaggttt 950
||||||| ||||||||| ||||||||| | ||||||||| ||| |||||||||
901 acgaacgcaaccatctaccTGAatcaatttttacatataaggttt 950

951 agtttatcggtgtta.caagtcacacattgtgtcggtgtägttaattcaa 999
||||||| ||||||||| ||| ||||||||| ||| ||||| ||||| |||||
951 agtttatcggtgttataaaatcacacatccgtatcggttttagtaagtcaa 1000

1000 agttaccatactctttttagaattttttttagtatacggtcgccggag 1049
||||| ||||| ||| | | | ||||||||| |||||||||
1001 agttaagatacttccttcttagaataattttttagtatacggtcgccggat 1050

12/16

Figure 9 (cont.)

```
1050 ttacggttacaaaggccaaatctattgtgtggaattccattattaaaaaa 1099
          ||||| |||||           |     | ||||||||||||||||| ||||| |||||
1051 atactgttac.....actattcggtgtggaattccattataaaaaaa 1091
          .
          .
          .
1100 taaaaaattagagtttgttagtttatctggtgatcaatatcaatataatt 1149
          |||||  |||
1092 ataaaaaaaaaaaaaaaaaaaaaa
```

13/16

Figure 10: Gap of SEQ ID NO: 2 and SEQ ID NO: 4

1 AISVFSTSYSFHKNLLLHSKQDILNRPCLLFSPVVVESPMRKKKTHRAAC 50
||||||. |||:||||| || .|| ||||||||: ||||||| | | |
1 AISVFSSGYSFYKNLLLDSPNILKPPCLLFSPVVIMSPMRKKKHDPC 50

51 ICSVVAERTRNLDIPQIEEEEENEELIEQTDSGIIHIKKTLGGKQSRRST 100
||||| ||||||||||||| ||||||| |:||||||| |||:
51 ICSVAGRTRNLDIPQIEEEEENVEELIEQTDSDIVHIKKTLGGKQSKRPT 100

101 GSIVAPVSLGILSMIGPAVYFKFSRLMECGDIPVAEMGITFAAFVAAAI 150
||||||| ||||||| ||||||| ||||| ||||||| ||||||| |||||:
101 GSIVAPVSLGILSMIGPAVYFKFSRLMEGGDIPVAEMGITFATFVAAAV 150

151 GTEFLSGWVHKELWHDSDLWYIHKSHHRSRKGRFEFNDVFAIINALPAIAL 200
||||| |||||:||||| ||||||| ||||||| ||||||| |||||||
151 GTEFLSAWVHKELWHESLWYIHKSHHRSRKGRFEFNDVFAIINALPAIAL 200

201 INYGFSNEGLLPGACFGTGLGTTVCGMAYIFLHNGLSHRRFPVGLIANVP 250
||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||
201 INYGFSNEGLLPGACFGVGLGTTVCGMAYIFLHNGLSHRRFPVWLIANVP 250

251 YFHKLAAAHQIHHSGKFQGVPGFLGPQELEEVRGGTEELERVISRTAK 300
||||||| ||||||| ||||||| ||||||| .||||||| ||||||| |
251 YFHKLAAAHQIHHSGKFQGVPGFLGPKELEEVRGGTEELERVISRTTK 300

301 RTQSST* 307
||| |||
301 RTQPST* 307

14/16

Figure 11: Gap of SEQ ID NO: 2 and Arabidopsis β -carotene hydroxylase (SEQ ID NO: 10)

```

1 AISVFSTSYSFHKNLLLHSKQDILNRPCLLFSPVVVESPMRKKKTHRAAC 50
     . ||. .|    ||    . | | . : | .
1 MAAXLSTAVTFKP...LHRSFSSSSDFRLRLPKSLSGFSPSLRFKRF SV 47
     .
51 ICSVVAERTRNLDIPQIEEEEENEELIEQTDGIIHIKKTLGGKQSRRST 100
     | || .| | | | | : : : . | | .| || |
48 CYVVEERRQNSPIENDERPESTSSTNAIDAELYLALRLAEKLERKKSERST 97
     .
101 GSIVAPVSCLGILSMIGPAVYFKFSRLMECGDIPVAEMGITFAAFVAAAI 150
     || .| || || || |::|| | | | | : | || | || | || |:
98 YLIAAMLSSFGITSMAVMAYYRFWSQMEGGEISMLEMGTFALSVGA AV 147
     .
151 GTEFLSGWVHKELWHDSLWYIHKSHRSRKGRFEFNDVFAIINALPAIAL 200
     | || .| | | : || | || .| .| | | : | . | | | | | | | |
148 GMEFWARWAHRALWHASLWNMHESHHKPREGPFELNDVFAIVNAGPAIGL 197
     .
201 INYGFSNEGLLPGACFGTGLGTTVCGMAYIFLHNGLSHRRFPVGLIANVP 250
     : .| | | | .| | | | | | | | | | | | | | | | | | | | | |
198 LSYGFFNKGLVPGLCFGAGLGITVFGIAYMFVHDGLVHKRFPVGP IADVP 247
     .
251 YFHKLAAAHQIHHSGKFQGVPGFLGPQELEEVRGGTEELERVISRTAK 300
     | | .| | | | : | | . | | | | : | | | | | | | | | | | | |
248 YLRKVAAAHQLHHTDKFNGVPYGLFLGPKELEEV.GGNEELDKEISRRIK 296
     .
301 RTQSST*..... 307
     .
297 SYKKASGSGSSSS* 311

```

15/16

Figure 12A (SEQ ID NO: 11)

1 CATAACCATAA ATAGTAGAGG ACAACCTACA AACCAACCAC CAGAAACCTC 50
51 CAATGGCAGC

Figure 12B (SEQ ID NO: 12)

MAAAISVFSSGYSFYKQNLLLDSKPNILKPPCLLFSPVVIMS PMRKXXKHGDPCICSVAGR
TRNLDIPOIEEEENVEELIEQTDSDIVHKKTLGGKQSKRPTGSIVAPVSCLGILSMIG
PAVYFKFSRLMEGGDIPVAEMGITFATFVAAVGTEFLSAWVHKELWHESLWYIHKSHHR
SRKGRFEFNDVFAIINALPAIALINYGFSNEGLLPGACFGVGLGTTVCGMAYIFLHNGLS
HRRFPVWLIANVPYFHKLAAAHQIHHSGKFQGVFFGLFLGPKELEEVVRGGTEELERVISR
TTKRTQPST*

16/16

Figure 13

SEQUENCE LISTING

<110> CUNNINGHAM, Francis X.

<120> CAROTENOID KETOLASE GENES AND GENE PRODUCTS, PRODUCTION
OF KETOCAROTENOIDS AND METHODS OF MODIFYING CAROTENOIDS
USING THE GENES

<130> 8172-9022

<140> Unknown

<141> 1999-05-21

<150> 60/086,460

<151> 1998-05-22

<160> 18

<170> PatentIn Ver. 2.0

<210> 1

<211> 1176

<212> DNA

<213> Adonis aestivalis

<400> 1

agcaatctca gtgttcagta caagttattc tttccacaag aatctttgt tgcaactaaa 60
acaagacatt ctcaaccgccc catgtttgct cttctctcca gttgtgtgg agtcgcctat 120
gagaaagaaa aagacacatc gtgctgcatt tatctgtct gttgcagaga gaacaaggaa 180
ccttgatatt cctcaaatgg aagaagagga agagaacgag gaagaactaa tagaacagac 240
ggattctggc ataattcata taaagaaaaac gctagggggg aaacaatcaa gacggtccac 300
tggctccatt gtgcaccccg tatcttgct tggatcctt tcaatgatcg gacctgctgt 360
ttacttcaag tttcacggc taatggatg tggagatatt cctgtgcag aaatggggat 420
tacgtttgcc gcctttgtgg ctgctgcatt tggcacggaa tttttgtcag gatgggttca 480
caaagaactc tggcacgatc ctttggta cattcacaag ttcacccata ggtcacgaaa 540
aggccgccttc gagttcaatg atgtgttgc tattattaaac ggcgttcctg ctattgtct 600
tatcaattat ggattctcaa atgaaggcct cttccctgga gcctgtttt gtaccggct 660
tggaaacgaca gtctgtggca tggcttacat ttttcttacat aatggcctt cacaccgaaag 720
gttcccagta gggcttattg caaacgtccc ttatccac aagctggctg cagctcacca 780
aatccatcac tcaggaaat ttcagggtgt accatttggc ctgttcctt gaccccgagga 840
attggaagaa gtaagaggag gcactgaaga attggagagg gtgatcagtc gtacagctaa 900
acgaacgcaa tcatttacat gaatcaactc ttttacattt atgaggtttt agtttatcg 960
tgttacaagt cacacatttgc tgcgttgta gtaattcaaa gttaccatac tcttttttag 1020
aattttttt tgatgtatag gtcgcggagt tacggttaca aaggccaaat ctattgtgt 1080
ggaattccat tattaaaaat aaaaattaga gttttagttt ttatctggtg atcaatatca 1140
atataatatta attaaagcaa aaaaaaaaaa aaaaaaa 1176

<210> 2

<211> 306

<212> PRT

<213> Adonis aestivalis

<400> 2

Ala Ile Ser Val Phe Ser Thr Ser Tyr Ser Phe His Lys Asn Leu Leu

1 5 10 15

Leu His Ser Lys Gln Asp Ile Leu Asn Arg Pro Cys Leu Leu Phe Ser
20 25 30Pro Val Val Val Glu Ser Pro Met Arg Lys Lys Lys Thr His Arg Ala
35 40 45Ala Cys Ile Cys Ser Val Ala Glu Arg Thr Arg Asn Leu Asp Ile Pro
50 55 60Gln Ile Glu Glu Glu Glu Asn Glu Glu Glu Leu Ile Glu Gln Thr
65 70 75 80Asp Ser Gly Ile Ile His Ile Lys Lys Thr Leu Gly Gly Lys Gln Ser
85 90 95Arg Arg Ser Thr Gly Ser Ile Val Ala Pro Val Ser Cys Leu Gly Ile
100 105 110Leu Ser Met Ile Gly Pro Ala Val Tyr Phe Lys Phe Ser Arg Leu Met
115 120 125Glu Cys Gly Asp Ile Pro Val Ala Glu Met Gly Ile Thr Phe Ala Ala
130 135 140Phe Val Ala Ala Ala Ile Gly Thr Glu Phe Leu Ser Gly Trp Val His
145 150 155 160Lys Glu Leu Trp His Asp Ser Leu Trp Tyr Ile His Lys Ser His His
165 170 175Arg Ser Arg Lys Gly Arg Phe Glu Phe Asn Asp Val Phe Ala Ile Ile
180 185 190Asn Ala Leu Pro Ala Ile Ala Leu Ile Asn Tyr Gly Phe Ser Asn Glu
195 200 205Gly Leu Leu Pro Gly Ala Cys Phe Gly Thr Gly Leu Gly Thr Thr Val
210 215 220

Cys Gly Met Ala Tyr Ile Phe Leu His Asn Gly Leu Ser His Arg Arg

225	230	235	240
Phe Pro Val Gly Leu Ile Ala Asn Val Pro Tyr Phe His Lys Leu Ala			
245	250	255	
Ala Ala His Gln Ile His His Ser Gly Lys Phe Gln Gly Val Pro Phe			
260	265	270	
Gly Leu Phe Leu Gly Pro Gln Glu Leu Glu Glu Val Arg Gly Gly Thr			
275	280	285	
Glu Glu Leu Glu Arg Val Ile Ser Arg Thr Ala Lys Arg Thr Gln Ser			
290	295	300	
Ser Thr			
305			

<210> 3
 <211> 1112
 <212> DNA
 <213> Adonis aestivalis

<400> 3
 agcaattca gtgttcagtt caggattttc tttctacaag aatctttgt tggactcaaa 60
 accaaatatt ctcaaacccc catgcctgct attctctcca gttgtatca tgcgcctat 120
 gagaaagaaa aagaaacatg gtgatccatg tatctgctcc gttgcaggaa gaacaaggaa 180
 cttgtatatt cctcaaattt aagaagagga agagaatgtg gaagaactaa tagaacagac 240
 cgattctgac atagtgacata taaagaaaac actagggggg aaacaatcaa aacggcccac 300
 tggctccatt gtcgcacccg tatcttgtct tggatcctt tcaatgattt gacctgctgt 360
 ttacttcaag tttcacggc taatggaggg tggagatata cctgtacgaa aatggggat 420
 tacgttgcc acctttgtt ctgctgtgt tggcacggag ttttgcacatg catgggttca 480
 caaagaactc tggcacgagt ctttgttgc cattcacaag ttcaccatc ggtcacgaaa 540
 agggcccttc gagttcaatg atgtgttgc tattttaac gcgcttcccg ctattgctct 600
 tatcaattat ggattctcca atgaaggcct cttccttgc gcgtgccttgc tggcggtct 660
 tggacaacaaca gtctgttgc tggcttacat ttttcttacat aatggcttat cacaccgaag 720
 gttcccaatg tggcttattt cgaacgtccc ttatttccac aagctggctg cagctcacca 780
 aatacaccac tcaggaaaat ttcagggtgt accatttggc ctgttcccttgc gacccaagga 840
 attggaaagaa gtaagaggag gcactgaaga gttggagagg gtaatcagtc gtacaactaa 900
 acgaacgcaa ccatctaccat gaatcaattt ttttacatata ataaggttt agtttatcg 960
 tggatataaaa tcacacatcc gtatcggtt agtaagtcaa agttaagata ctcccttctt 1020
 agaatatttt ttgtatgtata ggtcgccgat atactgttac actattcggtt gtggaaattcc 1080
 attataaaaaa aataaaaaaa aaaaaaaaaaa aa 1112

<210> 4
 <211> 306
 <212> PRT
 <213> Adonis aestivalis

<400> 4

Ala Ile Ser Val Phe Ser Ser Gly Tyr Ser Phe Tyr Lys Asn Leu Leu
1 5 10 15

Leu Asp Ser Lys Pro Asn Ile Leu Lys Pro Pro Cys Leu Leu Phe Ser
20 25 30

Pro Val Val Ile Met Ser Pro Met Arg Lys Lys Lys His Gly Asp
35 40 45

Pro Cys Ile Cys Ser Val Ala Gly Arg Thr Arg Asn Leu Asp Ile Pro
50 55 60

Gln Ile Glu Glu Glu Glu Asn Val Glu Glu Leu Ile Glu Gln Thr
65 70 75 80

Asp Ser Asp Ile Val His Ile Lys Lys Thr Leu Gly Gly Lys Gln Ser
85 90 95

Lys Arg Pro Thr Gly Ser Ile Val Ala Pro Val Ser Cys Leu Gly Ile
100 105 110

Leu Ser Met Ile Gly Pro Ala Val Tyr Phe Lys Phe Ser Arg Leu Met
115 120 125

Glu Gly Gly Asp Ile Pro Val Ala Glu Met Gly Ile Thr Phe Ala Thr
130 135 140

Phe Val Ala Ala Ala Val Gly Thr Glu Phe Leu Ser Ala Trp Val His
145 150 155 160

Lys Glu Leu Trp His Glu Ser Leu Trp Tyr Ile His Lys Ser His His
165 170 175

Arg Ser Arg Lys Gly Arg Phe Glu Phe Asn Asp Val Phe Ala Ile Ile
180 185 190

Asn Ala Leu Pro Ala Ile Ala Leu Ile Asn Tyr Gly Phe Ser Asn Glu
195 200 205

Gly Leu Leu Pro Gly Ala Cys Phe Gly Val Gly Leu Gly Thr Thr Val
210 215 220

Cys Gly Met Ala Tyr Ile Phe Leu His Asn Gly Leu Ser His Arg Arg
225 230 235 240

Phe Pro Val Trp Leu Ile Ala Asn Val Pro Tyr Phe His Lys Leu Ala

245

250

255

Ala Ala His Gln Ile His His Ser Gly Lys Phe Gln Gly Val Pro Phe
260 265 270

Gly Leu Phe Leu Gly Pro Lys Glu Leu Glu Glu Val Arg Gly Gly Thr
275 280 285

Glu Glu Leu Glu Arg Val Ile Ser Arg Thr Thr Lys Arg Thr Gln Pro
290 295 300

Ser Thr
305

<210> 5
<211> 1205
<212> DNA
<213> Adonis aestivalis

<400> 5
gggctgcagg aattcggcac gagagaatac tcagtgttca gtacaagtta ttctttccac 60
aagaatctct tggcactc aaaacaagac attctcaacc gcccattttt gctcttctct 120
ccagttgtgg tggagtcgac tatgagaaag aaaaagacac atcgtgctgc atgtatctgc 180
tctgttgcag agagaacaag gaacctgtat attcctcaaa ttgaagaaga ggaagagaac 240
gaggaagaac taatagaaca gacggattct ggcataattc atataaagaa aacgctaggg 300
gggaaacaat caagacggc cactggctcc attgtcgac ccgtatcttgc tcttgggatc 360
ctttcaatga tcggacctgc tggttacttc aagttttcac ggctaatggc gtgtggagat 420
attcctgtcg cagaaatggc gattacgtt gcccctttt tggctgctgc gattggcacg 480
gaattttgtt caggatggc tcacaaagaa ctctggcacg attctttgtt gtacattcac 540
aagtctcacc ataggcacgc aaaaggccgc ttcgagttca atgatgttt tgctattttt 600
aacgcgccttc ctgttattgc tcttataat tatggattct caaatgaagg cctccttctt 660
ggagcctgct ttggatccgc tcttggaaacg acagtctgtg gcatggctta catttttctt 720
cacaatggcc ttccacaccgc aaggttccca gttagggctta ttgcaaacgt cccttatttc 780
cacaagctgg ctgcagtc ccaaattccat cactcaggaa aatttcagggt tgtaaccattt 840
ggcctgttcc ttggacccca ggaattggaa gaagtaagag gaggcactga agaattggag 900
agggtgatca gtcgtacacgc taaacgaacg caatcatcta catgaatcaa ctctttaca 960
tttatgaggt tttagtttgc cggtgttaca agtcacacat ttgtgtcggtt gttagtaattt 1020
aaagttacca tactttttt tagaattttt ttttgatgtt taggtcgccg agttacgggtt 1080
acaaaaggcca aatctattgc tggtggattt cattattaaa aataaaaaatt agagttgtt 1140
gttttatctg gtgatcaatc tcaatataata ttaattaaag caaaaaaaaaa aaaaaaaaaac 1200
tcgag 1205

<210> 6
<211> 315
<212> PRT
<213> Adonis aestivalis

<400> 6

Met Gly Leu Gln Glu Phe Gly Thr Arg Ala Ile Ser Val Phe Ser Thr
1 5 10 15Ser Tyr Ser Phe His Lys Asn Leu Leu Leu His Ser Lys Gln Asp Ile
20 25 30Leu Asn Arg Pro Cys Leu Leu Phe Ser Pro Val Val Val Glu Ser Pro
35 40 45Met Arg Lys Lys Lys Thr His Arg Ala Ala Cys Ile Cys Ser Val Ala
50 55 60Glu Arg Thr Arg Asn Leu Asp Ile Pro Gln Ile Glu Glu Glu Glu
65 70 75 80Asn Glu Glu Glu Leu Ile Glu Gln Thr Asp Ser Gly Ile Ile His Ile
85 90 95Lys Lys Thr Leu Gly Gly Lys Gln Ser Arg Arg Ser Thr Gly Ser Ile
100 105 110Val Ala Pro Val Ser Cys Leu Gly Ile Leu Ser Met Ile Gly Pro Ala
115 120 125Val Tyr Phe Lys Phe Ser Arg Leu Met Glu Cys Gly Asp Ile Pro Val
130 135 140Ala Glu Met Gly Ile Thr Phe Ala Ala Phe Val Ala Ala Ala Ile Gly
145 150 155 160Thr Glu Phe Leu Ser Gly Trp Val His Lys Glu Leu Trp His Asp Ser
165 170 175Leu Trp Tyr Ile His Lys Ser His His Arg Ser Arg Lys Gly Arg Phe
180 185 190Glu Phe Asn Asp Val Phe Ala Ile Ile Asn Ala Leu Pro Ala Ile Ala
195 200 205Leu Ile Asn Tyr Gly Phe Ser Asn Glu Gly Leu Leu Pro Gly Ala Cys
210 215 220Phe Gly Thr Gly Leu Gly Thr Thr Val Cys Gly Met Ala Tyr Ile Phe
225 230 235 240Leu His Asn Gly Leu Ser His Arg Arg Phe Pro Val Gly Leu Ile Ala
245 250 255

Asn Val Pro Tyr Phe His Lys Leu Ala Ala Ala His Gln Ile His His
260 265 270

Ser Gly Lys Phe Gln Gly Val Pro Phe Gly Leu Phe Leu Gly Pro Gln
275 280 285

Glu Leu Glu Glu Val Arg Gly Gly Thr Glu Glu Leu Glu Arg Val Ile
290 295 300

Ser Arg Thr Ala Lys Arg Thr Gln Ser Ser Thr
305 310 315

<210> 7

<211> 1141

<212> DNA

<213> Adonis aestivalis

<400> 7

gggctgcagg aattcggcac gagagcaatt tcagtgtca gttcaggta ttctttctac 60
aagaatctct tggactc aaaaccaaatttctcaaac ccccatgcct gctattctct 120
ccagttgtga tcatgtcgcc tatgagaaaag aaaaagaaaac atggtgatcc atgtatctgc 180
tccgttgcag ggagaacaag gaaccttgat attcctcaaa ttgaagaaga ggaagagaat 240
gtgaaagaac taatagaaca gaccgattct gacatagtgc atataaagaa aacacttaggg 300
gggaaacaat caaaacggcc cactggctcc attgtcgac ccgtatcttg tcttgggatc 360
ctttcaatga ttggacctgc tgtttacttc aagttttcac ggctaatttgc gggtgagat 420
atacctgttag cagaaatggg gattacgttt gccacctttg ttgctgctgc tggcgcacg 480
gagttttgtt cagcatgggt tcacaaaagaa ctctggcacg agtctttgtt gtacatttcac 540
aagtctcacc atcggtcacg aaaaggccgc ttcgagttca atgatgtgtt tgcttattt 600
aacgcgttcc cgcgttccatgc tcttataat tatggattct ccaatgaagg cctccctt 660
ggagcgtgct ttgggtgcg tcttggaaaca acagtctgtt gtatggctta catttttttt 720
cacaatggcc tatcacaccg aaggttccca gtatggctta ttgcgaacgt cccttatttc 780
cacaagctgg ctgcagctca ccaaatacac cactcaggaa aatttcagggt tgcaccattt 840
ggcctgttcc ttggacccaa ggaattggaa gaagtaagag gaggcactga agagttggag 900
aggtaatca gtcgtacaac taaacgaacg caaccatcta cctgaatcaa tttttttaca 960
tatataaggt tttagtttat cggtgtata aaatcacaca tccgtatcgt tttagtaagt 1020
caaagttaag atacttcattt cttttagatat tttttagatgt ataggtcgac gatatactgt 1080
tacactattt gttgtggaaat tccattataa aaaaataaaaa aaaaaaaaaaaa aaaaactcgaa 1140
g 1141

<210> 8

<211> 315

<212> PRT

<213> Adonis aestivalis

<400> 8

Met Gly Leu Gln Glu Phe Gly Thr Arg Ala Ile Ser Val Phe Ser Ser

1

5

10

15

Gly Tyr Ser Phe Tyr Lys Asn Leu Leu Leu Asp Ser Lys Pro Asn Ile
20 25 30

Leu Lys Pro Pro Cys Leu Leu Phe Ser Pro Val Val Ile Met Ser Pro
35 40 45

Met Arg Lys Lys Lys His Gly Asp Pro Cys Ile Cys Ser Val Ala
50 55 60

Gly Arg Thr Arg Asn Leu Asp Ile Pro Gln Ile Glu Glu Glu Glu
65 70 75 80

Asn Val Glu Glu Leu Ile Glu Gln Thr Asp Ser Asp Ile Val His Ile
85 90 95

Lys Lys Thr Leu Gly Gly Lys Gln Ser Lys Arg Pro Thr Gly Ser Ile
100 105 110

Val Ala Pro Val Ser Cys Leu Gly Ile Leu Ser Met Ile Gly Pro Ala
115 120 125

Val Tyr Phe Lys Phe Ser Arg Leu Met Glu Gly Gly Asp Ile Pro Val
130 135 140

Ala Glu Met Gly Ile Thr Phe Ala Thr Phe Val Ala Ala Ala Val Gly
145 150 155 160

Thr Glu Phe Leu Ser Ala Trp Val His Lys Glu Leu Trp His Glu Ser
165 170 175

Leu Trp Tyr Ile His Lys Ser His His Arg Ser Arg Lys Gly Arg Phe
180 185 190

Glu Phe Asn Asp Val Phe Ala Ile Ile Asn Ala Leu Pro Ala Ile Ala
195 200 205

Leu Ile Asn Tyr Gly Phe Ser Asn Glu Gly Leu Leu Pro Gly Ala Cys
210 215 220

Phe Gly Val Gly Leu Gly Thr Thr Val Cys Gly Met Ala Tyr Ile Phe
225 230 235 240

Leu His Asn Gly Leu Ser His Arg Arg Phe Pro Val Trp Leu Ile Ala
245 250 255

Asn Val Pro Tyr Phe His Lys Leu Ala Ala Ala His Gln Ile His His

260

265

270

Ser Gly Lys Phe Gln Gln Val Pro Phe Gly Leu Phe Leu Gly Pro Lys
 275 280 285

Glu Leu Glu Glu Val Arg Gly Gly Thr Glu Glu Leu Glu Arg Val Ile
 290 295 300

Ser Arg Thr Thr Lys Arg Thr Gln Pro Ser Thr
 305 310 315

<210> 9

<211> 1149

<212> DNA

<213> Adonis aestivalis

<400> 9

agcaatctca gtgttcagtt caagttattc tttccacaag aatctttgt tgcactcaaa 60
 acaagacatt ctcacccccc catgtttgtctt cttctctcca gttgtgggtgg agtcgcctat 120
 gagaaaagaaa aagacacatc gtgctgcatt tatctgtctt gttgcagaga gaacaaggaa 180
 ctttgcattt cctcaatattt aagaagagga agagaacgag gaagaactaa tagaacagac 240
 ggattctggc ataattcata taaagaaaac gctagggggg aaacaatcaa gacggtccac 300
 tggctccatt gtcgcacccg tatcttgcatt tggatcctt tcaatgatcg gacctgctgt 360
 ttacttcaag ttttacggc taatggatgt tggagatatt cctgtgcag aaatggggat 420
 tacgtttgcc gcctttgttg ctgctgcgtt tggcacggaa tttttgtcag gatgggttca 480
 caaagaactc tggcacgtt ctttgcgtt cattcacaag tctcaccata ggtcacgaaa 540
 aggccgccttc gagttcaatc atgtgtttgc tattattaaac gcgcttcctt ctattgctct 600
 tatcaattat ggattctcaa atgaaggcct ctttccttgc gcctgtttt gtaccgggtct 660
 tggAACGACA gtctgtggca tggcttacat ttttcttgcatt aatggcctt cacaccgaa 720
 gttcccagta gggcttattt caaacgtccc ttatccac aagctgctg cagctcacca 780
 aatccatcac tcaggaaaat ttcagggtgt accatttggc ctgttccttgc gaccccaggaa 840
 atttggaaat gtaagaggaa gcaactgaaga atttggagagg gtgatcagtc gtacagctaa 900
 acgaacgcaa tcatactacat gaatcaactc ttttacattt atgaggtttt agtttatcgg 960
 tggatcaatcgtt cacaatattt tggctgttgcatt gtaattcaaa gttaccatac tcttttttag 1020
 aatttttttt tggatgtataat gtcgcggagt tacggttaca aaggccaaat ctattgttgc 1080
 ggaattccat tattaaaaat aaaaattaga gtttgcgttgcattt atcaatatca 1140
 atatatatt 1149

<210> 10

<211> 310

<212> PRT

<213> Arabidopsis

<400> 10

Met Ala Ala Xaa Leu Ser Thr Ala Val Thr Phe Lys Pro Leu His Arg
 1 5 10 15

Ser Phe Ser Ser Ser Ser Thr Asp Phe Arg Leu Arg Leu Pro Lys Ser
20 25 30

Leu Ser Gly Phe Ser Pro Ser Leu Arg Phe Lys Arg Phe Ser Val Cys
35 40 45

Tyr Val Val Glu Glu Arg Arg Gln Asn Ser Pro Ile Glu Asn Asp Glu
50 55 60

Arg Pro Glu Ser Thr Ser Ser Thr Asn Ala Ile Asp Ala Glu Tyr Leu
65 70 75 80

Ala Leu Arg Leu Ala Glu Lys Leu Glu Arg Lys Lys Ser Glu Arg Ser
85 90 95

Thr Tyr Leu Ile Ala Ala Met Leu Ser Ser Phe Gly Ile Thr Ser Met
100 105 110

Ala Val Met Ala Val Tyr Tyr Arg Phe Ser Trp Gln Met Glu Gly Gly
115 120 125

Glu Ile Ser Met Leu Glu Met Phe Gly Thr Phe Ala Leu Ser Val Gly
130 135 140

Ala Ala Val Gly Met Glu Phe Trp Ala Arg Trp Ala His Arg Ala Leu
145 150 155 160

Trp His Ala Ser Leu Trp Asn Met His Glu Ser His His Lys Pro Arg
165 170 175

Glu Gly Pro Phe Glu Leu Asn Asp Val Phe Ala Ile Val Asn Ala Gly
180 185 190

Pro Ala Ile Gly Leu Leu Ser Tyr Gly Phe Phe Asn Lys Gly Leu Val
195 200 205

Pro Gly Leu Cys Phe Gly Ala Gly Leu Gly Ile Thr Val Phe Gly Ile
210 215 220

Ala Tyr Met Phe Val His Asp Gly Leu Val His Lys Arg Phe Pro Val
225 230 235 240

Gly Pro Ile Ala Asp Val Pro Tyr Leu Arg Lys Val Ala Ala His
245 250 255

Gln Leu His His Thr Asp Lys Phe Asn Gly Val Pro Tyr Gly Leu Phe
260 265 270

Leu Gly Pro Lys Glu Leu Glu Glu Val Gly Gly Asn Glu Glu Leu Asp
275 280 285

Lys Glu Ile Ser Arg Arg Ile Lys Ser Tyr Lys Lys Ala Ser Gly Ser
290 295 300

Gly Ser Ser Ser Ser Ser
305 310

<210> 11
<211> 60
<212> DNA
<213> Adonis aestivalis

<400> 11
cataccataa atagtagagg acaacctaca aaccaaccac cagaaacctc caatggcagc 60

<210> 12
<211> 309
<212> PRT
<213> Adonis aestivalis

<400> 12
Met Ala Ala Ala Ile Ser Val Phe Ser Ser Gly Tyr Ser Phe Tyr Lys
1 5 10 15

Asn Leu Leu Leu Asp Ser Lys Pro Asn Ile Leu Lys Pro Pro Cys Leu
20 25 30

Leu Phe Ser Pro Val Val Ile Met Ser Pro Met Arg Lys Lys Lys Lys
35 40 45

His Gly Asp Pro Cys Ile Cys Ser Val Ala Gly Arg Thr Arg Asn Leu
50 55 60

Asp Ile Pro Gln Ile Glu Glu Glu Glu Asn Val Glu Glu Leu Ile
65 70 75 80

Glu Gln Thr Asp Ser Asp Ile Val His Ile Lys Lys Thr Leu Gly Gly
85 90 95

Lys Gln Ser Lys Arg Pro Thr Gly Ser Ile Val Ala Pro Val Ser Cys
100 105 110

Leu Gly Ile Leu Ser Met Ile Gly Pro Ala Val Tyr Phe Lys Phe Ser
115 120 125

Arg Leu Met Glu Gly Gly Asp Ile Pro Val Ala Glu Met Gly Ile Thr
130 135 140

Phe Ala Thr Phe Val Ala Ala Ala Val Gly Thr Glu Phe Leu Ser Ala
145 150 155 160

Trp Val His Lys Glu Leu Trp His Glu Ser Leu Trp Tyr Ile His Lys
165 170 175

Ser His His Arg Ser Arg Lys Gly Arg Phe Glu Phe Asn Asp Val Phe
180 185 190

Ala Ile Ile Asn Ala Leu Pro Ala Ile Ala Leu Ile Asn Tyr Gly Phe
195 200 205

Ser Asn Glu Gly Leu Leu Pro Gly Ala Cys Phe Gly Val Gly Leu Gly
210 215 220

Thr Thr Val Cys Gly Met Ala Tyr Ile Phe Leu His Asn Gly Leu Ser
225 230 235 240

His Arg Arg Phe Pro Val Trp Leu Ile Ala Asn Val Pro Tyr Phe His
245 250 255

Lys Leu Ala Ala Ala His Gln Ile His His Ser Gly Lys Phe Gln Gly
260 265 270

Val Pro Phe Gly Leu Phe Leu Gly Pro Lys Glu Leu Glu Glu Val Arg
275 280 285

Gly Gly Thr Glu Glu Leu Glu Arg Val Ile Ser Arg Thr Thr Lys Arg
290 295 300

Thr Gln Pro Ser Thr
305

<210> 13

<211> 310

<212> PRT

<213> Arabidopsis

<400> 13

Met Ala Ala Xaa Leu Ser Thr Ala Val Thr Phe Lys Pro Leu His Arg
1 5 10 15

Ser Phe Ser Ser Ser Ser Thr Asp Phe Arg Leu Arg Leu Pro Lys Ser
20 25 30

Leu Ser Gly Phe Ser Pro Ser Leu Arg Phe Lys Arg Phe Ser Val Cys
35 40 45

Tyr Val Val Glu Glu Arg Arg Gln Asn Ser Pro Ile Glu Asn Asp Glu
50 55 60

Arg Pro Glu Ser Thr Ser Ser Thr Asn Ala Ile Asp Ala Glu Tyr Leu
65 70 75 80

Ala Leu Arg Leu Ala Glu Lys Leu Glu Arg Lys Lys Ser Glu Arg Ser
85 90 95

Thr Tyr Leu Ile Ala Ala Met Leu Ser Ser Phe Gly Ile Thr Ser Met
100 105 110

Ala Val Met Ala Val Tyr Tyr Arg Phe Ser Trp Gln Met Glu Gly Gly
115 120 125

Glu Ile Ser Met Leu Glu Met Phe Gly Thr Phe Ala Leu Ser Val Gly
130 135 140

Ala Ala Val Gly Met Glu Phe Trp Ala Arg Trp Ala His Arg Ala Leu
145 150 155 160

Trp His Ala Ser Leu Trp Asn Met His Glu Ser His His Lys Pro Arg
165 170 175

Glu Gly Pro Phe Glu Leu Asn Asp Val Phe Ala Ile Val Asn Ala Gly
180 185 190

Pro Ala Ile Gly Leu Leu Ser Tyr Gly Phe Phe Asn Lys Gly Leu Val
195 200 205

Pro Gly Leu Cys Phe Gly Ala Gly Leu Gly Ile Thr Val Phe Gly Ile
210 215 220

Ala Tyr Met Phe Val His Asp Gly Leu Val His Lys Arg Phe Pro Val
225 230 235 240

Gly Pro Ile Ala Asp Val Pro Tyr Leu Arg Lys Val Ala Ala His
245 250 255

Gln Leu His His Thr Asp Lys Phe Asn Gly Val Pro Tyr Gly Leu Phe
260 265 270

Leu Gly Pro Lys Glu Leu Glu Glu Val Gly Gly Asn Glu Glu Leu Asp
275 280 285

Lys Glu Ile Ser Arg Arg Ile Lys Ser Tyr Lys Lys Ala Ser Gly Ser
290 295 300

Gly Ser Ser Ser Ser Ser
305 310

<210> 14
<211> 305
<212> PRT
<213> Arabidopsis

<400> 14
Met Ala Ala Gly Leu Ser Thr Ile Ala Val Thr Leu Lys Pro Leu Asn
1 5 10 15

Arg Ser Ser Phe Ser Ala Asn His Pro Ile Ser Thr Ala Val Phe Pro
20 25 30

Pro Ser Leu Arg Phe Asn Gly Phe Arg Arg Arg Lys Ile Leu Thr Val
35 40 45

Cys Phe Val Val Glu Glu Arg Lys Gln Ser Ser Pro Met Asp Asp Asp
50 55 60

Asn Lys Pro Glu Ser Thr Thr Ser Ser Ser Glu Ile Leu Met Thr Ser
65 70 75 80

Arg Leu Leu Lys Lys Ala Glu Lys Lys Ser Glu Arg Phe Thr Tyr
85 90 95

Leu Ile Ala Ala Val Met Ser Ser Phe Gly Ile Thr Ser Met Ala Ile
100 105 110

Met Ala Val Tyr Tyr Arg Phe Ser Trp Gln Met Lys Gly Gly Glu Val
115 120 125

Ser Val Leu Glu Met Phe Gly Thr Phe Ala Leu Ser Val Gly Ala Ala
130 135 140

Val Val Gly Met Glu Phe Trp Ala Arg Trp Ala His Arg Ala Leu Trp
145 150 155 160

His Asp Ser Leu Trp Asn Met His Glu Ser His His Lys Pro Arg Glu
165 170 175

Gly Ala Phe Glu Leu Asn Asp Val Phe Ala Ile Thr Asn Ala Val Pro

180	185	190
Ala Ile Gly Leu Leu Tyr Tyr Gly Phe Leu Asn Lys Gly Leu Val Pro		
195	200	205
Gly Leu Cys Phe Gly Ala Gly Leu Gly Ile Thr Met Phe Gly Met Ala		
210	215	220
Tyr Met Phe Val His Asp Gly Leu Val His Lys Arg Phe Pro Val Gly		
225	230	235
Pro Ile Ala Asn Val Pro Tyr Leu Arg Lys Val Ala Ala Ala His Gln		
245	250	255
Leu His His Thr Asp Lys Phe Lys Gly Val Pro Tyr Gly Leu Phe Leu		
260	265	270
Gly Pro Lys Gln Glu Val Glu Val Gly Gly Lys Glu Glu Leu Glu		
275	280	285
Lys Glu Ile Ser Arg Arg Ile Lys Leu Tyr Asn Lys Gly Ser Ser Thr		
290	295	300
Ser		
305		

<210> 15
 <211> 315
 <212> PRT
 <213> Capsicum annuum

<400> 15			
Met Ala Ala Glu Ile Ser Ile Ser Ala Ser Ser Arg Ala Ile Cys Leu			
1	5	10	15
Gln Arg Asn Pro Phe Pro Ala Pro Lys Tyr Phe Ala Thr Ala Pro Pro			
20	25	30	
Leu Leu Phe Phe Ser Pro Leu Thr Cys Asn Leu Asp Ala Ile Leu Arg			
35	40	45	
Ser Arg Arg Lys Pro Arg Leu Ala Ala Cys Phe Val Leu Lys Asp Asp			
50	55	60	
Lys Leu Tyr Thr Ala Gln Ser Gly Lys Gln Ser Asp Thr Glu Ala Ile			
65	70	75	80

Gly Asp Glu Ile Glu Val Glu Thr Asn Glu Glu Lys Ser Leu Ala Val
85 90 95

Arg Leu Ala Glu Lys Phe Ala Arg Lys Lys Ser Glu Arg Phe Thr Tyr
100 105 110

Leu Val Ala Ala Val Met Ser Ser Leu Gly Ile Thr Ser Met Ala Val
115 120 125

Ile Ser Val Tyr Tyr Arg Phe Ser Trp Gln Met Glu Gly Gly Glu Met
130 135 140

Pro Phe Ser Glu Met Phe Cys Thr Phe Ala Leu Ala Phe Gly Ala Ala
145 150 155 160

Ile Gly Met Glu Tyr Trp Ala Arg Trp Ala His Arg Ala Leu Trp His
165 170 175

Ala Ser Leu Trp His Met His Glu Ser His His Arg Pro Arg Glu Gly
180 185 190

Pro Phe Glu Leu Asn Asp Ile Phe Ala Ile Ile Asn Ala Val Pro Ala
195 200 205

Ile Ala Phe Phe Ser Phe Gly Phe Asn His Lys Gly Leu Ile Pro Gly
210 215 220

Ile Cys Phe Gly Ala Gly Leu Gly Ile Thr Val Phe Gly Met Ala Tyr
225 230 235 240

Met Phe Val His Asp Gly Leu Val His Lys Arg Phe Pro Val Gly Pro
245 250 255

Ile Ala Lys Val Pro Tyr Phe Gln Arg Val Ala Ala Ala His Gln Leu
260 265 270

His His Ser Asp Lys Phe Asp Gly Val Pro Tyr Gly Leu Phe Leu Gly
275 280 285

Pro Lys Glu Leu Glu Glu Val Gly Val Ile Glu Glu Leu Glu Lys Glu
290 295 300

Val Asn Arg Arg Ile Lys Ser Leu Lys Arg Leu
305 310 315

<210> 16

<211> 316

<212> PRT

<213> Capsicum annuum

<400> 16

Thr Thr Gly Arg Tyr His Tyr Gln Leu Val Trp Cys Gln Ile Ser Phe
1 5 10 15Ser Ser Thr Ser Arg Thr Ser Tyr Tyr Arg His Ser Pro Phe Leu Gly
20 25 30Pro Lys Pro Thr Pro Thr Pro Ser Val Tyr Pro Ile Thr Pro Phe
35 40 45Ser Pro Asn Leu Gly Ser Ile Leu Arg Cys Arg Arg Arg Pro Ser Phe
50 55 60Thr Val Cys Phe Val Leu Glu Asp Asp Lys Phe Lys Thr Gln Phe Glu
65 70 75 80Ala Gly Glu Glu Asp Ile Glu Met Lys Ile Glu Glu Gln Ile Ser Ala
85 90 95Thr Arg Leu Ala Glu Lys Leu Ala Arg Lys Lys Ser Glu Arg Phe Thr
100 105 110Tyr Leu Val Ala Ala Val Met Ser Ser Phe Gly Ile Thr Ser Met Ala
115 120 125Val Met Ala Val Tyr Tyr Arg Phe Tyr Trp Gln Met Glu Gly Gly Glu
130 135 140Val Pro Phe Ser Glu Met Phe Gly Thr Phe Ala Leu Ser Val Gly Ala
145 150 155 160Ala Val Gly Met Glu Phe Trp Ala Arg Trp Ala His Lys Ala Leu Trp
165 170 175His Ala Ser Leu Trp His Met His Glu Ser His His Lys Pro Arg Glu
180 185 190Gly Pro Phe Glu Leu Asn Asp Val Phe Ala Ile Ile Asn Ala Val Pro
195 200 205Ala Ile Ala Leu Leu Asp Tyr Gly Phe Phe His Lys Gly Leu Ile Pro
210 215 220Gly Leu Cys Phe Gly Ala Gly Leu Gly Ile Thr Val Phe Gly Met Ala
225 230 235 240

Tyr Met Phe Val His Asp Gly Leu Val His Lys Arg Phe Pro Val Gly
245 250 255

Pro Val Ala Asn Val Pro Tyr Leu Arg Lys Val Ala Ala Ala His Ser
260 265 270

Leu His His Ser Glu Lys Phe Asn Gly Val Pro Tyr Gly Leu Phe Leu
275 280 285

Gly Pro Lys Glu Leu Glu Glu Val Gly Gly Leu Glu Glu Leu Glu Lys
290 295 300

Glu Val Asn Arg Arg Thr Arg Tyr Ile Lys Gly Ser
305 310 315

<210> 17

<211> 29

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic

<400> 17

cagaatcggt ctgttctatt agttcttcc

29

<210> 18

<211> 32

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic

<400> 18

caatttgagg aatatcaagg ttccttgttc tc

32

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US99/10455

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) :C12P 23/00, 7/26; C12N 9/02, 1/20, 15/00; C07H 21/04; C07K 14/00
US CL :435/67, 148, 189, 252.3, 252.33, 320.1; 536/23.2, 23.6; 530/350

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/67, 148, 189, 252.3, 252.33, 320.1; 536/23.2, 23.6; 530/350

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Please See Extra Sheet.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 5,453,565 A (MAWSON) 26 September 1995, see abstract and claims.	1-20
A,E	US 5,910,433 A (KAIJIWARA et al.) 08 June 1999, see the entire patent.	1-20
Y,P	US 5,811,273 A (MISAWA et al.) 22 September 1998, See abstract, column 30 - lines 48-58 and claims.	1-20

Further documents are listed in the continuation of Box C.

See patent family annex.

• Special categories of cited documents:	• T	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
• A* document defining the general state of the art which is not considered to be of particular relevance	• X*	document of particular relevance: the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
• E* earlier document published on or after the international filing date	• Y*	document of particular relevance: the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
• L* document which may throw doubt on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)		
• O* document referring to an oral disclosure, use, exhibition or other means	• Z*	document member of the same patent family
• P* document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search	Date of mailing of the international search report
13 AUGUST 1999	29 OCT 1999

Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer TEKCHAND SAIDHA Telephone No. (703) 308-0196
-------------------------------------------------------------------------------------------------------------------------------------------------------	-----------------------------------------------------------------------

Form PCT/ISA/210 (second sheet)(July 1992)*

INTERNATIONAL SEARCH REPORT

International application No. PCT/US99/10455

B. FIELDS SEARCHED

Electronic data bases consulted (Name of data base and where practicable terms used):

APS, STN Files: Medline, Caplus, Biosis, Agricola, Embase & Scisearch. Search terms used : beta carotene and ketolase, ketocarotenoid, Adonis aestivalis, carotenoid biosynthesis, gene? or dna or rna or nucleic acid? in various permutations and combinations.